

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 1, 2002, 14:59:10 ; Search time 24.84 seconds

(without alignments)
377.193 Million cell updates/sec

Title: US-09-591-632-2_COPY_1_123

Perfect score: 701

Sequence: 1 MSDSNOGNNQNTQYQSQNG.....FNVNNNLQGYQAGFQPSQG 123

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	701	100.0	685	1	EFBY52
2	166	23.7	405	2	S19355
3	166	23.7	741	2	S12921
4	163.5	23.3	589	2	S71954
5	148	21.1	419	2	T04886
6	147.5	21.0	409	2	T20847
7	147.5	21.0	945	2	T21998
8	147	21.0	462	1	LUD07
9	144	20.5	232	2	T21735
10	143	20.4	372	2	T64223
11	142	20.3	676	2	S41022
12	141	20.1	404	2	S46269
13	141	20.1	800	2	T51653
14	140.5	20.0	236	2	T18240
15	139	19.8	632	2	T02627
16	138	19.7	1111	2	T28070
17	137.5	19.6	284	2	T23158
18	137.5	19.6	755	2	T15447
19	137	19.5	462	2	C29349
20	136.5	19.5	461	2	T10265
21	136	19.4	462	4	S33798
22	136	19.4	526	1	S33799
23	135	19.3	128	2	S45871
24	135	19.3	528	2	G02127
25	135	19.3	867	2	T27136
26	135	19.3	871	2	T27135
27	133	19.0	3498	2	T22330
28	132	18.8	882	2	H70199
29	131	18.7	1161	2	T18400

ALIGNMENTS

30	128	18.3	618	2	T19407	hypothetical prote
31	128	18.3	735	2	T19403	hypothetical prote
32	126.5	18.0	638	2	S37876	glutamine-rich pro
33	124.5	17.8	382	2	T29339	hypothetical prote
34	124.5	17.8	1196	2	S65245	translation elonga
35	123	17.5	346	2	S44874	ZC21.3 protein - C
36	121.5	17.3	353	1	S56750	single stranded D
37	121	17.3	597	2	T51889	related to clathrin
38	121	17.3	954	1	S20907	endo-1,4-beta-xyla
39	120.5	17.2	989	2	T02568	hypothetical prote
40	119.5	17.0	301	2	JM0079	heterogeneous nucl
41	119.5	17.0	557	2	T19686	hypothetical prote
42	119.5	17.0	1139	2	T33275	hypothetical prote
43	118.5	16.9	129	2	T06256	dormancy-associate
44	118	16.8	592	2	T21536	hypothetical prote
45	118	16.8	838	1	EWTHW	glutennin, high mol

RESULT 1

EFBY52
Suppressor 2 protein - yeast (Saccharomyces cerevisiae)
N:Alternate names: G1-to-S transition protein; protein YD9395.05; protein YDR172w
C:Species: Saccharomyces cerevisiae
C:Date: 31-Dec-1991 #sequence revision 31-Dec-1991 #extl change 19-Jan-2001
C:Accession: S00733; J03323; S49768; S00488; A26742; S00533; S05723
R:Wilson, P.G.; Gilbertson, M.R.
J. Mol. Biol. 199, 559-573, 1988
A:Title: SUP2 suppressor protein of yeast. A fusion protein related to the EF-1 faml
A:Reference number: S00733; MUID:88172503
A:Accession: S00733
A:Molecule type: DNA
A:Residues: 1-685 <MIL>
A:Cross-references: EMBL:X07163; NID:94581; PIDN:CAA30155.1; PID:94582
R:Kushnir, V.V.; Ter-Avanesyan, M.D.; Telckov, M.V.; Surguchov, A.P.; Smirnov, V.N.
Gene 66, 45-54, 1988
A:Title: Nucleotide sequence of the SUP2(SUP35) gene of Saccharomyces cerevisiae.
A:Reference number: J03323; MUID:88329727
A:Accession: J03323
A:Molecule type: DNA
A:Residues: 1-685 <KUS>
A:Cross-references: EMBL:M21129; NID:g172789; PIDN:AAA3513.1; PID:g172791
R:Murphy, L.; Harris, D.E.
Submitted to the EMBL Data Library, November 1994
A:Reference number: S49764
A:Accession: S49768
A:Molecule type: DNA
A:Residues: 1-685 <MUR>
A:Cross-references: EMBL:Z46727; NID:g1289283; PIDN:CAA6677.1; PID:g1289287; GSPDB:G
R:Kikuchi, Y.; Shimatake, H.; Kikuchi, A.
EMBO J. 7, 1175-1182, 1988
A:Title: A yeast gene required for the G1-to-S transition encodes a protein containin
A:Reference number: S00488; MUID:88296422
A:Accession: S00488
A:Molecule type: DNA
A:Residues: 1-52, 'C', 54-685 <KIK>
A:Cross-references: GB:Y00829; EMBL:Y00859; NID:g3711; PIDN:CAA68760.1; PID:g3712
C:Genetics:
A:Gene: SCD: SUP35; SUP12; GST1; SUP2; MIPS:YDR172w
A:Cross-references: SCD: S0002579; MIPS:YDR172w
A:Map position: 4R
C:Superfamily: suppressor 2 protein; translation elongation factor Tu homology
C:Keywords: duplication; GTP binding; nucleotide binding; P-loop; phosphoprotein; tan
F:1-123/Domain: A <DOM1>
F:42-119/Region: 10-residue repeats
F:124-253/Domain: charged <DOM2>
F:159-222/Region: glutamic acid/lysine-rich
F:254-685/Domain: C <DOM4>
F:261-409/Domain: translation elongation factor Tu homology <ETU>
F:267-274/Region: nucleotide-binding motif A (P-loop)
F:406-409/Region: GTP-binding NKXD motif

A:Reference number: A40977; MUID:91373375
A:Accession: A40977

```

Db      25  GMYVQYQSQSQSYYPSSGGQ---YRG--GYGNSFGMGQGYGNGQDSTVQDFQYQGYQR  79
OY      57  -QGRTQQTNPDAQYQQQTINPGGTQD-----LYNQG-----  86
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      80  GGGGGRGYQ---GYGNSNRSGSYPERDSSMYSFMDLTNYGYSQGYQYGYSSASY  136
OY      87  -----GY---QOQFNPQGGRGNYKFNFYNNNLGCGYQAGFQFQSQS  123
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      137 QMTPTVQNONQCTMQQQYNTPT-SQSSMSPTMTGNGYQSQSQSQQSSMG  185

RESULT  10
164223
bifunctional endo-1,4-beta-xyylanase homolog - Mycoplasma genitalium
C:Species: Mycoplasma genitalium
C:Date: 17-Nov-1995 #sequence_revision 17-Nov-1995 #text_change 07-Dec-1999
C:Accession: I64223
R:Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R
M.; Fuhmann, J.; Nguyen, D.; Uterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick
, C.A.; Venter, J.C.
Science 270, 397-403, 1995
A:Title: The minimal gene complement of Mycoplasma genitalium.
A:Reference number: A64200; MUID:96026346
A:Accession: I64223
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-372 <TIG>

```

C:Genetics :
A:Genetic code : SGC3

Query Match	20.4%	Score 143	DB 2	Length 372
Best Local Similarity	32.7%	Pred No. 1.5e-05		
Matches 51; Conservative	12;	Mismatches 55;	Indels 38;	Gaps 8

```

QY      3  DSNQGNQNTQQYISQNG---NQDQGN-----RYQYQAYNAQAQAPAGGYQNTQYQYISG  54
      || |:::|| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      22  DQNGYQDPPNQDQFNQSGFDPNQDQFNQPGFDPNQYQYQDPPNQDQFNQAGFQDQNGYQD  81

```

```
OY      55 -----YQGGGQQYNPDAGY-QQQYNPQGCGYYQYNPQGCGYQQQF--NP 94
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      82 PNOQGFNQPGFDPNQGYQDPNQQQFN-QAGFDQNQYQDPPNQQQGNQSGFGDNQYVQDP 140
```

```

Qy  95 QGGRGNKKNENYNNNLQYQAGFQ-----PSQ 122
      :  :  :  :  :  :  :  :
Db 141 NQQGFNQPSFDLNNQ-QFNQPSGFNQSPAFEITPEQ 175

```

RESULT	11
S41022	

C:Species: Caenorhabditis elegans
C:date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 09-Sep-1995
C:Accession: S41022

submitted to the EMBL Data Library, January 1994
A:Reference number: S41014
A:Accession: S41022
A:Status: Released

A: Molecule type: DNA
A: Residues: 1-676 <BER>
A: Cross-References: EMBL:Z29443; NID:g1067051; PID:g443836

A:introns: 69/3; 161/1; 208/3; 227/1; 357/1; 505/3
C:Superfamily: annexin repeat homology
F:373-444/Domain: annexin repeat homology <AX1>
D:145-173/Domain: annexin repeat homology <AX2>

F:604-675/Domain: annexin <AXX>

A:Reference number: Z14681
 A:Accession: T02627
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-632 <R0U>
 A:Cross-references: EMBL:AC004747; NID:g3413696; PID:g3413716
 A:Experimental source: cultivar Columbia
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Bentto, M.I.; Town, C.D.; Fujii, C.Y.;
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
 Neufuss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
 Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A84420; MUID:20083487
 A:Accession: H84654
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-632 <STO>
 A:Cross-references: GB:AE002093; NID:g3413716; PIDN:MAC31239.1; GSPDB:GN00139
 C:Genetics:
 A:Gene: T19L18.23; At2g25960
 A:Map position: 2
 A:Introns: 112/1; 206/3; 240/3; 264/3; 304/3
 C:Superfamily: glutenin

Query Match 19.8%; Score 139; DB 2; Length 632;
 Best local Similarity 31.7%; Pred. No.5.8e-05;
 Matches 46; Conservative 6; Mismatches 53; Indels 40; Gaps 7;

QY	13	YQQYSONGNOQGNRRYQGYQAYNAQAQAPAG-----YYQNYQGY-----SGYQ	56
Db	374	YDQSSVPPSQSAQGEYDYGGQOQSQOPSSGSSAPPTDTTGYNYQYQHASGYGQAGGYQ	433
QY	57	QGGYQOQYNPDAGYQOQYNPDGQGYQGYNPDGQGYQOQGFNP-----QGGRGNY	101
Db	434	QDGYGAYN--ASQOQSGYGAAGYDQ--QGGYSTTNPSQEDASQAAPPSSAQSGQAGY	488
QY	102	KNFNYYNNLQGY--QAGF--QPOSQ	122
Db	489	GTTGQOPRPAQSGTQAGYGAPTSQ	513

Search completed: February 1, 2002, 15:02:47
 Job time: 217 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 1, 2002, 14:58:20 ; Search time 19.74 Seconds
(without alignments)
140.218 Million cell updates/sec

Title: US-09-591-632-2_COPY_1_123
Perfect score: 701
Sequence: 1 MSDSNGNNGNQYQYQSQNG.....FNYYNNLQGYQGFQPOSOG 123

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCrUS_COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfillsl.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	121	17.3	485	2	US-08-749-391-2
2	121	17.3	485	4	US-09-390-200-2
3	111.5	15.9	296	2	US-08-700-637-4
4	104	14.8	307	1	US-07-982-112-2
5	103.5	14.8	362	1	US-08-437-027-21
6	103.5	14.8	365	1	US-08-437-027-20
7	103.5	14.8	656	2	US-08-343-448-2
8	103.5	14.8	656	4	US-09-214-564A-4
9	103	14.7	341	2	US-08-538-711A-8
10	103	14.7	341	4	US-08-725-027-8
11	103	14.7	353	2	US-08-538-711A-7
12	103	14.7	353	4	US-08-725-027-7
13	93	13.3	542	1	US-07-814-964-13
14	93	13.3	542	1	US-08-258-442-13
15	93	13.3	542	1	US-08-328-809-8
16	93	13.3	542	3	PCr-US92-11107-13
17	91	13.0	1160	3	US-08-808-599A-24
18	90	12.8	493	4	US-08-556-978B-59
19	90	12.8	529	4	US-09-247-806-2
20	90	12.8	729	4	US-09-625-188-20
21	89.5	12.8	1895	2	US-08-619-554-4
22	89.5	12.8	430	2	US-08-945-848-8
23	89.5	12.8	806	1	US-07-980-528-2
24	89	12.7	303	4	US-08-818-112-92
25	87	12.4	405	2	US-08-282-197C-61
26	86	12.3	678	5	PCr-US93-03027-3
27	85	12.1	595	1	US-08-425-069-4

28	85	12.1	595	2	US-08-317-844B-4	Sequence 4, Appl1
29	84	12.0	545	2	US-08-990-114-1	Sequence 1, Appl1
30	83.5	11.9	655	4	US-09-347-833-4	Sequence 4, Appl1
31	81.5	11.6	660	3	US-09-058-489-18	Sequence 18, Appl1
32	81.5	11.6	660	3	US-09-058-489-91	Sequence 91, Appl1
33	81.5	11.6	1876	2	US-08-619-554-2	Sequence 2, Appl1
34	79	11.3	714	4	US-08-556-978B-61	Sequence 61, Appl1
35	79	11.3	714	4	US-09-247-806-10	Sequence 10, Appl1
36	78.5	11.2	724	3	US-08-793-331-4	Sequence 6, Appl1
37	78.5	11.2	725	3	US-08-793-331-6	Sequence 6, Appl1
38	77.5	11.1	314	2	US-08-525-742-6	Sequence 6, Appl1
39	77.5	11.1	788	2	US-08-918-914-4	Sequence 4, Appl1
40	76.5	10.9	234	4	US-09-146-770-4	Sequence 4, Appl1
41	76.5	10.9	234	4	US-09-216-295-4	Sequence 4, Appl1
42	76	10.8	227	1	US-08-458-023B-4	Sequence 4, Appl1
43	76	10.8	396	1	US-07-702-771-1	Sequence 1, Appl1
44	76	10.8	396	1	US-08-438-883-1	Sequence 1, Appl1
45	76	10.8	396	1	US-08-741-599-1	Sequence 1, Appl1

ALIGNMENTS

RESULT 1
US-08-749-391-2
Sequence 2, Application US/08749391
Patent No. 5948667
GENERAL INFORMATION:
APPLICANT: Cheng, Kuo-Joan
APPLICANT: Sellinger, Leonard B.
APPLICANT: Liu, Jin-Hao
APPLICANT: Hu, Youji
APPLICANT: Forsberg, Cecil W.
APPLICANT: Moloney, Maurice W.
TITLE OF INVENTION: A Xylanase Obtained From an
NUMBER OF SEQUENCES: 6
TITLE OF INVENTION: Anaerobic Fungus
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
STREET: 5370 Manhattan Circle, Suite 201
CITY: Boulder
STATE: Colorado
COUNTRY: USA
ZIP: 80803
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/749,391
FILING DATE: 13-NOV-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Donna M. Feiber
REGISTRATION NUMBER: 33,878
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 499-8080
TELEFAX: (303) 499-8089
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 485 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-749-391-2

Query Match 17.3% Score 121, DB 2; Length 485;
Best Local Similarity 35.5% Pred. No. 0.00013;
Matches 49; Conservative 8; Mismatches 55; Indels 26; Gaps 8;

Query Match	14.8%;	Score 103.5;	DB 1;	Length 365;
Best Local Similarity	31.8%;	Pred. No. 0.0052;		
Matches 35; Conservative	4;	Mismatches 48;	Indels 23;	Gaps 4;

QY 72 QYNQGSYYQY-----NPGGYYQOQPNQGGKRNKNNYNNNLGQYQAG 116
 Db 5 DYSYSSQAANQ-----QGSAYT - AQPGYVAQQTQAAGQGSYGTYYQPIIDYVYTA 55
 QY 12 NYOQSYONGNQOQGNRRGYQAYAAQAPAGGYQYNNQYSGYQSGGYQYYPDAGYQO 71
 Db 5 QTTATTCQTATATSYGQPTGGTTPPAQA-----TSQPVQSGTGG 96

RESULT 7
US-08-343-443B-2

GENERAL INFORMATION:
APPLICANT: Auriat, Alain
APPLICANT: Delattre, Olivier
APPLICANT: Desmaze, Chantal
APPLICANT: Melot, Thomas
APPLICANT: Peter, Martine
APPLICANT: Ploougastel, Beatrice
APPLICANT: Thomas, Gilles
APPLICANT: Zuchan, Tessica
TITLE OF INVENTION: NUCLEIC ACID CORRESPONDING TO A GENE OF
TITLE OF INVENTION: CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL
TITLE OF INVENTION: TRANSLATIONS ASSOCIATED WITH THE DEVELOPMENT OF CANCEROUS
TITLE OF INVENTION: TUMORS, AND NUCLEIC ACIDS OF FUSION RESULTING FROM SAID
TITLE OF INVENTION: TRANSLATIONS
NUMBER OF SEQUENCES: 129

Query Match 14.8%; Score 103.5; DB 2; Length 656;

Best Local Similarity 31.88; Pred. No. 0.011;
Matches 35; Conservative 4; Mismatches 48; Indels 23; Gaps 4;

Oy 12 NVOOYSSQNGNOCOGNNRRVGVGYAYAAQAQPAGGYYQONTGVSGYGQQGGVQYQYNPAGVYQO 71
| | | | | | | | | | | | | | | | :
Dd 5 DYSTSQAAQAQ-----QGTSATV--AQPTGVYAQTGTAVGCQGSYGRVGCQDTVDSTPA 55
| | | | | | | | | | | | | | | | :
Oy 72 QYNPGGCVQY----NPQGVQYQGFENPGGKGNKNFNYNNNLCGYQAG 116
| | | | | | | | | | | | | | | | :
Dd 56 QTATVVGOTATAVTSGCPPTGTTFTAPA-----YSOPVGGTIGTG 96
| | | | | | | | | | | | | | | | :

RESULT 8
US-09-214-564A-4
; Sequence 4, Application US/09214564A

```

1  GENERAL INFORMATION:
2  APPLICANT: Sharp, Phillip A.
3  APPLICANT: Zhou, Qiang
4  TITLE OF INVENTION: Tat-sf: Cofactor For Stimulation Of Transcriptional
5  TITLE OF INVENTION: Elongation By HIV-1 TAT
6  TITLE OF INVENTION:
7  FILE REFERENCE: M0656/7042
8  CURRENT APPLICATION NUMBER: US/09/214,564A
9  CURRENT FILING DATE: 1999-08-18
10 PRIOR APPLICATION NUMBER: US 60/021,218
11 PRIOR FILING DATE: 1996-07-03
12 PRIOR APPLICATION NUMBER: US 60/033,152
13 PRIOR FILING DATE: 1996-12-13
14 PRIOR APPLICATION NUMBER: PCT/US97/11713
15 PRIOR FILING DATE: 1997-07-03
16 NUMBER OF SEQ ID NOS: 5
17 SOFTWARE: FastSeq for Windows Version 3.0
18 SEQ ID NO 4
19 LENGTH: 656
20 TYPE: PRT
21 ORGANISM: Homo sapiens
22 US-09-214-564A-4

```

Query Match	14.88;	Score 103.5;	DB 4;	Length 656;
Best Local Similarity	31.88;	Pred. No. 0.011;		
Matches 35; Conservative	4;	Mismatches 48;	Indels 23;	Gaps 4

[illegible]

RESULT 9
US-08-538-711A-8
Sequence 8, Application US/08538711A
Patent No. 5994062
GENERAL INFORMATION:
APPLICANT: MULSHINE, JAMES, L.
TITLE OF INVENTION: AN EPITHELIAL PROTEIN AND
TITLE OF INVENTION: DNA THEREOF FOR USE IN EARLY CANCER DETECTION
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII

GenCore version 4.5
Copyright (c) 1993 - 2000 Compen Ltd.

OM protein - protein search, using sw model

Run on: February 1, 2002, 14:57:35 ; Search time 38.77 Seconds
(without alignments)
235.002 Million cell updates/sec

Title: US-09-591-632-2_COPY_1_123
Perfect score: 701
Sequence: 1 MSNSQGNNOONQYQSYQNG.....FNNNNLQGYQAFQPOPOSG 123

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A.Geneseq_1101:*

1: /SID8/gcgdata/geneseq/geneseq/AA1980.DAT:*

2: /SID8/gcgdata/geneseq/geneseq/AA1981.DAT:*

3: /SID8/gcgdata/geneseq/geneseq/AA1982.DAT:*

4: /SID8/gcgdata/geneseq/geneseq/AA1983.DAT:*

5: /SID8/gcgdata/geneseq/geneseq/AA1984.DAT:*

6: /SID8/gcgdata/geneseq/geneseq/AA1985.DAT:*

7: /SID8/gcgdata/geneseq/geneseq/AA1986.DAT:*

8: /SID8/gcgdata/geneseq/geneseq/AA1987.DAT:*

9: /SID8/gcgdata/geneseq/geneseq/AA1988.DAT:*

10: /SID8/gcgdata/geneseq/geneseq/AA1989.DAT:*

11: /SID8/gcgdata/geneseq/geneseq/AA1990.DAT:*

12: /SID8/gcgdata/geneseq/geneseq/AA1991.DAT:*

13: /SID8/gcgdata/geneseq/geneseq/AA1992.DAT:*

14: /SID8/gcgdata/geneseq/geneseq/AA1993.DAT:*

15: /SID8/gcgdata/geneseq/geneseq/AA1994.DAT:*

16: /SID8/gcgdata/geneseq/geneseq/AA1995.DAT:*

17: /SID8/gcgdata/geneseq/geneseq/AA1996.DAT:*

18: /SID8/gcgdata/geneseq/geneseq/AA1997.DAT:*

19: /SID8/gcgdata/geneseq/geneseq/AA1998.DAT:*

20: /SID8/gcgdata/geneseq/geneseq/AA1999.DAT:*

21: /SID8/gcgdata/geneseq/geneseq/AA2000.DAT:*

22: /SID8/gcgdata/geneseq/geneseq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	701	100.0	685	22	AAB30792	Amino acid sequenc
2	682	97.3	271	22	AAB30800	A modified N regio
3	450	64.2	215	22	AAB30799	A modified N regio
4	228	32.5	715	22	AAB30820	Amino acid sequenc
5	166	23.7	405	22	AAB30810	Amino acid sequenc
6	166	23.7	405	22	AAB30821	Amino acid sequenc
7	166	23.7	741	22	AAB30819	Amino acid sequenc
8	160	22.8	407	10	AA94879	ORF incorporated w
9	148	21.1	417	21	AA630013	Arabidopsis thalia
10	148	21.1	419	21	AA630012	Arabidopsis thalia
11	148	21.1	439	21	AA630011	Arabidopsis thalia

12	141	20.1	404	17	AAR90675	Human chromosome-1
13	136.5	19.5	461	16	AAR75506	Nicotiana glauca ar
14	135	19.3	128	22	AAB30822	Amino acid sequenc
15	134	19.1	341	11	AAR06522	Tapetum-specific p
16	134	19.1	341	11	AAR09307	Male flower-specif
17	134	19.1	341	11	AAR09298	Male flower-specif
18	134	19.1	341	11	AAW10628	Anther-specific CD
19	132	18.8	437	16	AAR75507	Nicotiana glauca
20	129	18.4	341	11	AAR06451	PM510 contg. male
21	128	18.3	894	22	AAB35148	Human NFAR-2 SEQ I
22	121	17.3	280	21	AAB63197	Gene 21 human secr
23	121	17.3	485	19	AAW65462	Neocallimastix pat
24	121	17.3	485	20	AAV27283	N. patriciarum xyl
25	121	17.3	485	22	AAB35588	N. patriciarum end
26	121	17.3	954	21	AAB48550	Ruminococcus flave
27	120.5	17.2	989	21	AA633723	Arabidopsis thalia
28	120.5	17.2	994	21	AA633722	Arabidopsis thalia
29	120.5	17.2	1006	21	AA633721	Arabidopsis thalia
30	118.5	16.9	398	21	AA629375	Arabidopsis thalia
31	118.5	16.9	418	21	AA629374	Arabidopsis thalia
32	118.5	16.9	432	21	AA629373	Arabidopsis thalia
33	117	16.7	579	22	AAB83003	Mealworm encapsula
34	117	16.7	579	22	AAB83004	Mealworm encapsula
35	116	16.5	545	13	AA627493	p34 from M. paratu
36	115	16.4	103	21	AAV95064	Arabidopsis thalia
37	115	16.4	171	21	AA630911	Arabidopsis thalia
38	115	16.4	266	21	AA630910	Arabidopsis thalia
39	115	16.4	324	21	AA630909	Arabidopsis thalia
40	114	16.3	103	21	AAV95066	Arabidopsis thalia
41	111.5	15.9	946	21	AA643595	Arabidopsis thalia
42	111.5	15.9	972	21	AA643594	Arabidopsis thalia
43	111.5	15.9	1026	21	AA643593	Arabidopsis thalia
44	111	15.8	256	22	AAB30807	Amino acid sequenc
45	108.5	15.5	128	21	AAB58406	Lung cancer associ

ALIGNMENTS

RESULT	1
ID	AAB30792 standard; Protein: 685 AA.
XX	XX
AC	AAB30792:
XX	XX
DT	02-APR-2001 (first entry)
XX	XX
DE	Amino acid sequence of a yeast Sup35 protein
XX	XX
KW	Sup35; phenotype: SCHNG; self-coalesce; higher-order aggregate;
KW	amyloidogenic domain; aggregation; fibril; phenotypic alteration;
KW	gene therapy; disease resistance; plant pigmentation; prion disease.
XX	XX
OS	Saccharomyces cerevisiae.
XX	XX
PN	WO200075324-A2.
XX	XX
PD	14-DEC-2000.
XX	XX
PE	09-JUN-2000; 2000MO-US15876.
XX	XX
PR	09-JUN-1999; 99US-0138833.
XX	XX
PA	(ARCH-) ARCH DEV CORP.
XX	XX
PI	Lindquist S, Li L, Ma J, Liu J, Sondheimer N, Scheibel T;
DR	WPI: 2001-061723/07.
XX	XX
DR	N-PSDB; AAC6679.
XX	XX
PT	New nucleic acid encoding chimeric proteins with self-assembly
PT	properties, useful e.g. for diagnosis and treatment of prion diseases,
PT	also related aggregates, fibrils and polymers -

PR	14-MAY-1999:	990S-0134221.
PR	14-MAY-1999:	990S-0134370.
PR	18-MAY-1999:	990S-0134768.
PR	19-MAY-1999:	990S-0134941.
PR	20-MAY-1999:	990S-0135124.
PR	21-MAY-1999:	990S-0135353.
PR	24-MAY-1999:	990S-0135629.
PR	25-MAY-1999:	990S-0136021.
PR	27-MAY-1999:	990S-0136392.
PR	28-MAY-1999:	990S-0136782.
PR	01-JUN-1999:	990S-0137222.
PR	03-JUN-1999:	990S-0137528.
PR	04-JUN-1999:	990S-0137502.
PR	07-JUN-1999:	990S-0137724.
PR	08-JUN-1999:	990S-0138094.
PR	10-JUN-1999:	990S-0138540.
PR	10-JUN-1999:	990S-0138847.
PR	14-JUN-1999:	990S-0139119.
PR	16-JUN-1999:	990S-0139452.
PR	16-JUN-1999:	990S-0139453.
PR	17-JUN-1999:	990S-0139492.
PR	18-JUN-1999:	990S-0139454.
PR	18-JUN-1999:	990S-0139455.
PR	18-JUN-1999:	990S-0139456.
PR	18-JUN-1999:	990S-0139457.
PR	18-JUN-1999:	990S-0139458.
PR	18-JUN-1999:	990S-0139459.
PR	18-JUN-1999:	990S-0139460.
PR	18-JUN-1999:	990S-0139461.
PR	18-JUN-1999:	990S-0139462.
PR	18-JUN-1999:	990S-0139463.
PR	18-JUN-1999:	990S-0139750.
PR	21-JUN-1999:	990S-0139763.
PR	21-JUN-1999:	990S-0139817.
PR	22-JUN-1999:	990S-0139899.
PR	23-JUN-1999:	990S-0140353.
PR	23-JUN-1999:	990S-0140354.
PR	24-JUN-1999:	990S-0140695.
PR	28-JUN-1999:	990S-0140823.
PR	29-JUN-1999:	990S-0140991.
PR	30-JUN-1999:	990S-0141287.
PR	01-JUL-1999:	990S-0141842.
PR	01-JUL-1999:	990S-0142154.
PR	02-JUL-1999:	990S-0142055.
PR	06-JUL-1999:	990S-0142390.
PR	08-JUL-1999:	990S-0142803.
PR	09-JUL-1999:	990S-0142820.
PR	12-JUL-1999:	990S-0142977.
PR	13-JUL-1999:	990S-0143542.
PR	14-JUL-1999:	990S-0143624.
PR	15-JUL-1999:	990S-0144005.
PR	16-JUL-1999:	990S-0144085.
PR	16-JUL-1999:	990S-0144086.
PR	19-JUL-1999:	990S-0144325.
PR	19-JUL-1999:	990S-0144331.
PR	19-JUL-1999:	990S-0144332.
PR	19-JUL-1999:	990S-0144333.
PR	19-JUL-1999:	990S-0144334.
PR	19-JUL-1999:	990S-0144335.
PR	20-JUL-1999:	990S-0144352.
PR	20-JUL-1999:	990S-0144632.
PR	20-JUL-1999:	990S-0144884.
PR	21-JUL-1999:	990S-0144814.
PR	21-JUL-1999:	990S-0145086.
PR	21-JUL-1999:	990S-0145088.
PR	22-JUL-1999:	990S-0145085.
PR	22-JUL-1999:	990S-0145087.
PR	22-JUL-1999:	990S-0145089.
PR	22-JUL-1999:	990S-0145192.
PR	23-JUL-1999:	990S-0145145.
PR	23-JUL-1999:	990S-0145218.
PR	23-JUL-1999:	990S-0145224.
PR	26-JUL-1999:	990S-0145276.
<hr/>		
PR	27-JUL-1999:	990S-0145913.
PR	27-JUL-1999:	990S-0145918.
PR	27-JUL-1999:	990S-0145919.
PR	28-JUL-1999:	990S-0145951.
PR	02-AUG-1999:	990S-0146386.
PR	02-AUG-1999:	990S-0146388.
PR	02-AUG-1999:	990S-0146389.
PR	03-AUG-1999:	990S-0147038.
PR	04-AUG-1999:	990S-0147204.
PR	04-AUG-1999:	990S-0147302.
PR	05-AUG-1999:	990S-0147192.
PR	05-AUG-1999:	990S-0147260.
PR	06-AUG-1999:	990S-0147303.
PR	06-AUG-1999:	990S-0147416.
PR	09-AUG-1999:	990S-0147493.
PR	09-AUG-1999:	990S-0147935.
PR	10-AUG-1999:	990S-0148171.
PR	11-AUG-1999:	990S-0148319.
PR	12-AUG-1999:	990S-0148341.
PR	13-AUG-1999:	990S-0148565.
PR	13-AUG-1999:	990S-0148684.
PR	16-AUG-1999:	990S-0149368.
PR	17-AUG-1999:	990S-0149175.
PR	18-AUG-1999:	990S-0149426.
PR	20-AUG-1999:	990S-0149722.
PR	20-AUG-1999:	990S-0149723.
PR	20-AUG-1999:	990S-0149929.
PR	23-AUG-1999:	990S-0149902.
PR	23-AUG-1999:	990S-0149930.
PR	25-AUG-1999:	990S-0150566.
PR	26-AUG-1999:	990S-0150884.
PR	27-AUG-1999:	990S-0151065.
PR	27-AUG-1999:	990S-0151066.
PR	27-AUG-1999:	990S-0151080.
PR	30-AUG-1999:	990S-0151303.
PR	31-AUG-1999:	990S-0151338.
PR	01-SEP-1999:	990S-0151930.
PR	07-SEP-1999:	990S-0152363.
PR	10-SEP-1999:	990S-0153070.
PR	13-SEP-1999:	990S-0153758.
PR	15-SEP-1999:	990S-0154018.
PR	16-SEP-1999:	990S-0154039.
PR	20-SEP-1999:	990S-0154779.
PR	22-SEP-1999:	990S-0155139.
PR	23-SEP-1999:	990S-0155486.
PR	24-SEP-1999:	990S-0155659.
PR	28-SEP-1999:	990S-0156458.
PR	29-SEP-1999:	990S-0156596.
PR	04-OCT-1999:	990S-0157117.
PR	05-OCT-1999:	990S-0157753.
PR	06-OCT-1999:	990S-0157865.
PR	07-OCT-1999:	990S-0158029.
PR	08-OCT-1999:	990S-0158232.
PR	12-OCT-1999:	990S-0158369.
PR	13-OCT-1999:	990S-0159293.
PR	13-OCT-1999:	990S-0159294.
PR	13-OCT-1999:	990S-0159295.
PR	14-OCT-1999:	990S-0159329.
PR	14-OCT-1999:	990S-0159330.
PR	14-OCT-1999:	990S-0159331.
PR	14-OCT-1999:	990S-0159637.
PR	14-OCT-1999:	990S-0159638.
PR	18-OCT-1999:	990S-0159584.
PR	21-OCT-1999:	990S-0160741.
PR	21-OCT-1999:	990S-0160767.
PR	21-OCT-1999:	990S-0160768.
PR	21-OCT-1999:	990S-0160770.
PR	21-OCT-1999:	990S-0160814.
PR	21-OCT-1999:	990S-0160815.
PR	22-OCT-1999:	990S-0160980.
PR	22-OCT-1999:	990S-0160981.
PR	22-OCT-1999:	990S-0160989.
PR	25-OCT-1999:	990S-0161404.

PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145921.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 22-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 13-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158222.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.

PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 21.1%; Score 148; DB 21; Length 419;
Best Local Similarity 34.8%; Pred. No. 2.8e-07;
Matches 57; Conservative 9; Mismatches 42; Indels 56; Gaps 13;

QY 11 ONYQOYSQNGNO---QOCNNRY---QGY---QAYMAQAP--AGGYQN-----YQG 51
Db 215 gnfgrntbtygqppmqgsgyspqqyatlpgqgqgtatppfgygynqgprpppyq- 273
QY 52 YSGYQOG-----GYOQYN-----PDAGYQOOYVPOGXYQOYN--POGXYQO----- 90
Db 274 -agynqgqgspvpppyqagynqvgqspvpppygtqssyq-qqsgsnysagppqgynqgqpr 331
QY 91 QFNPOGGRGNYKNFNNNL-----QGYQAGFQPOSQ 122
Db 332 nypnqga-gnfgpasgagnlgpapqagupgygqgyspqqgqndq 374

RESULT 11
AAG30011
ID AAG30011 standard; Protein: 439 AA.
XX
AC AAG30011;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 35804.
XX
KW Protein identification: signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 9905-0134218.
PR 14-MAY-1999; 9905-0134219.
PR 14-MAY-1999; 9905-0134221.
PR 14-MAY-1999; 9905-0134370.
PR 18-MAY-1999; 9905-0134768.
PR 19-MAY-1999; 9905-0134941.
PR 20-MAY-1999; 9905-0135124.
PR 21-MAY-1999; 9905-0135353.
PR 24-MAY-1999; 9905-0135629.
PR 25-MAY-1999; 9905-0136021.
PR 27-MAY-1999; 9905-0136392.
PR 28-MAY-1999; 9905-0136782.
PR 01-JUN-1999; 9905-0137222.
PR 03-JUN-1999; 9905-0137528.
PR 04-JUN-1999; 9905-0137502.
PR 07-JUN-1999; 9905-0137724.
PR 08-JUN-1999; 9905-0138094.
PR 10-JUN-1999; 9905-0138540.
PR 10-JUN-1999; 9905-0138847.
PR 14-JUN-1999; 9905-0139119.
PR 16-JUN-1999; 9905-0139452.
PR 16-JUN-1999; 9905-0139453.
PR 17-JUN-1999; 9905-0139492.
PR 18-JUN-1999; 9905-0139454.
PR 18-JUN-1999; 9905-0139455.
PR 18-JUN-1999; 9905-0139456.
PR 18-JUN-1999; 9905-0139457.
PR 18-JUN-1999; 9905-0139458.
PR 18-JUN-1999; 9905-0139459.
PR 18-JUN-1999; 9905-0139460.
PR 18-JUN-1999; 9905-0139461.
PR 18-JUN-1999; 9905-0139462.
PR 18-JUN-1999; 9905-0139463.
PR 18-JUN-1999; 9905-0139750.
PR 18-JUN-1999; 9905-0139763.
PR 21-JUN-1999; 9905-0139817.
PR 22-JUN-1999; 9905-0139819.
PR 23-JUN-1999; 9905-0140353.
PR 23-JUN-1999; 9905-0140354.
PR 24-JUN-1999; 9905-0140695.
PR 28-JUN-1999; 9905-0140823.
PR 29-JUN-1999; 9905-0140991.
PR 30-JUN-1999; 9905-0141287.
PR 01-JUL-1999; 9905-0141842.
PR 01-JUL-1999; 9905-0142154.
PR 02-JUL-1999; 9905-0142055.
PR 06-JUL-1999; 9905-0142390.
PR 08-JUL-1999; 9905-0142803.
PR 09-JUL-1999; 9905-0142920.
PR 12-JUL-1999; 9905-0142977.
PR 13-JUL-1999; 9905-0143542.
PR 14-JUL-1999; 9905-0143624.
PR 15-JUL-1999; 9905-0144005.
PR 16-JUL-1999; 9905-0144085.
PR 16-JUL-1999; 9905-0144086.
PR 19-JUL-1999; 9905-0144325.
PR 19-JUL-1999; 9905-0144331.
PR 19-JUL-1999; 9905-0144332.
PR 19-JUL-1999; 9905-0144333.
PR 19-JUL-1999; 9905-0144334.
PR 19-JUL-1999; 9905-0144335.
PR 20-JUL-1999; 9905-0144352.
PR 20-JUL-1999; 9905-0144632.
PR 20-JUL-1999; 9905-0144884.
PR 21-JUL-1999; 9905-0144814.
PR 21-JUL-1999; 9905-0145086.
PR 21-JUL-1999; 9905-0145088.
PR 22-JUL-1999; 9905-0145085.
PR 22-JUL-1999; 9905-0145087.
PR 22-JUL-1999; 9905-0145089.
PR 22-JUL-1999; 9905-0145192.
PR 23-JUL-1999; 9905-0145145.
PR 23-JUL-1999; 9905-0145218.

PR 23-JUL-1999; 9905-0145224.
PR 26-JUL-1999; 9905-0145276.
PR 27-JUL-1999; 9905-01452913.
PR 27-JUL-1999; 9905-0145918.
PR 27-JUL-1999; 9905-0145919.
PR 28-JUL-1999; 9905-0145951.
PR 02-AUG-1999; 9905-0146386.
PR 02-AUG-1999; 9905-0146388.
PR 03-AUG-1999; 9905-0147038.
PR 04-AUG-1999; 9905-0147204.
PR 04-AUG-1999; 9905-0147302.
PR 05-AUG-1999; 9905-0147302.
PR 05-AUG-1999; 9905-0147260.
PR 06-AUG-1999; 9905-0147303.
PR 06-AUG-1999; 9905-0147416.
PR 09-AUG-1999; 9905-0147493.
PR 09-AUG-1999; 9905-0147935.
PR 10-AUG-1999; 9905-0148171.
PR 11-AUG-1999; 9905-0148319.
PR 12-AUG-1999; 9905-0148341.
PR 13-AUG-1999; 9905-0148365.
PR 13-AUG-1999; 9905-0148684.
PR 16-AUG-1999; 9905-0149368.
PR 17-AUG-1999; 9905-0149175.
PR 18-AUG-1999; 9905-0149426.
PR 20-AUG-1999; 9905-0149722.
PR 20-AUG-1999; 9905-0149723.
PR 20-AUG-1999; 9905-0149929.
PR 23-AUG-1999; 9905-0149902.
PR 23-AUG-1999; 9905-0149930.
PR 25-AUG-1999; 9905-0150566.
PR 26-AUG-1999; 9905-0150884.
PR 27-AUG-1999; 9905-0151065.
PR 27-AUG-1999; 9905-0151066.
PR 27-AUG-1999; 9905-0151080.
PR 30-AUG-1999; 9905-0151303.
PR 31-AUG-1999; 9905-0151438.
PR 01-SEP-1999; 9905-0151930.
PR 07-SEP-1999; 9905-0152363.
PR 10-SEP-1999; 9905-0153070.
PR 13-SEP-1999; 9905-0153758.
PR 15-SEP-1999; 9905-0154019.
PR 16-SEP-1999; 9905-0154039.
PR 20-SEP-1999; 9905-0154779.
PR 22-SEP-1999; 9905-0155139.
PR 23-SEP-1999; 9905-0155486.
PR 24-SEP-1999; 9905-0155659.
PR 28-SEP-1999; 9905-0156458.
PR 29-SEP-1999; 9905-0156596.
PR 04-OCT-1999; 9905-0157117.
PR 05-OCT-1999; 9905-0157753.
PR 06-OCT-1999; 9905-0157865.
PR 07-OCT-1999; 9905-0158029.
PR 08-OCT-1999; 9905-0158232.
PR 12-OCT-1999; 9905-0158369.
PR 13-OCT-1999; 9905-0159293.
PR 13-OCT-1999; 9905-0159294.
PR 13-OCT-1999; 9905-0159295.
PR 14-OCT-1999; 9905-0159329.
PR 14-OCT-1999; 9905-0159330.
PR 14-OCT-1999; 9905-0159331.
PR 14-OCT-1999; 9905-0159637.
PR 14-OCT-1999; 9905-0159638.
PR 18-OCT-1999; 9905-0159584.
PR 21-OCT-1999; 9905-0160741.
PR 21-OCT-1999; 9905-0160767.
PR 21-OCT-1999; 9905-0160768.
PR 21-OCT-1999; 9905-0160770.
PR 21-OCT-1999; 9905-0160814.
PR 21-OCT-1999; 9905-0160815.
PR 22-OCT-1999; 9905-0160980.
PR 22-OCT-1999; 9905-0160981.

DE	xx	
DE	xx	Tapetum-specific promoter encoded by clone pms10.
XX	xx	
KM	xx	Maize; sterility.
XX	xx	
PM	xx	WO9008831-A.
XX	xx	
PD	xx	09-AUG-1990.
XX	xx	
PF	xx	26-JAN-1990; 90WO-GB00114.
XX	xx	
PR	xx	26-JAN-1989; 89GB-0001675.
XX	xx	
PA	xx	(ICIL) IMPERIAL CHEM INDS PLC.
XX	xx	
PI	xx	Bridges IG, Bright SMU, Greenland AJ, Schuch WW, Reid GA;
XX	xx	
DR	xx	WPI: 1990-260940/34.
XX	xx	
DR	xx	N-PSDB: AAQ05759.
XX	xx	
PT	xx	Inhibition of gene expression in plants - which is useful in e.g.
PT	xx	impairing male sterility in maize.
XX	xx	
PS	xx	Disclosure: Fig 1; 40pp: English.
XX	xx	
CC	xx	DNA encoding the promoter is used in a gene construct with a dis-
CC	xx	rupter gene such as the mammalian uncoupling protein gene, or a
CC	xx	mutated form of the gene for the beta subunit of F1-ATPase. The
CC	xx	promoter operates in the target tissue and expression of the dis-
CC	xx	rupter gene inhibits respiration resulting in cell death. This is
CC	xx	of particular use for inhibiting male fertility in maize crops
CC	xx	where cell death results in failure to produce viable pollen. When
CC	xx	inserted into a gene cascade, as a module which permits external
CC	xx	control of expression, fertility may also be restored if required
CC	xx	See also AAR06524 and AAR06525.
XX	xx	
5Q	xx	Sequence 341 AA;

Query Match	Similarity	19.1%	Score 134	DB 11	Length 341
Best Local	Similarity	30.9%	Pred. No. 5.9e-06		
Matches	43	Conservative	11	Mismatches	35
				Indels	30
				Gaps	7
OY	2	SDSNOGNOANOYQYSONGNOGNNRYQGYAAYNAQAQAPAGYVO-----NYOGSYG	55		
Db	126	srpstrtdnrrnygrnygqgppqgn-----yg--nmrpppeggyqnmppqgnyqlyrsq	178		
OY	56	QOGGYOYNPDAGYVO-----QYNPGGYO-----QYNPGGAYQOENPGGRCNY	101		
Db	179	qdg--rgyapqgnyqagsgdqgrfgndyldtfsyngpdlftrsqyqghvnpagpaggqy	236		
OY	102	KNFNYYNNNL-OGYAGFQP	119		
Db	237	nmqgtrfsgqgqgqftr	255		

Search completed: February 1, 2002, 15:01:49
Job time: 254 sec

RESULT 2
 ID 020497 PRELIMINARY: PRT: 3498 AA.
 AC 020497;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1998 (TREMBLrel. 06, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE F47M4.2 PROTEIN.
 GN F47M4.2.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 NCBI_TaxID=6239;
 RP SEQUENCE FROM N.A.
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94150718; PubMed=7906398;
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 Craxton M., Dear S., Du Z., Durbin R., Favell A., Fulton L.,
 Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 Parsons J., Percy C., Ritken L., Roopra A., Saunders D., Showkhen R.,
 Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,
 Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 Watson A., Wellstock L., Wilkinson-Spoot J., Wohlman P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans.";
 RL Nature 368:32-38(1994).
 DR EMBL: 249888; CA90064.1; -
 DR InterPro: IPR000504; RRM.
 DR InterPro: IPR000561; EGF-like.
 DR PROSITE: PS00022; EGF_1; UNKNOWN_1.
 DR PROSITE: PS00030; RRM_RNP_1; UNKNOWN_1.
 SO SEQUENCE 3498 AA; 400778 MW; 283FEC65A6B9E336 CRC64;

Query Match 17.7%; Score 239; DB 5; Length 3498;
 Best Local Similarity 31.4%; Pred. No. 1e-12;
 Matches 101; Conservative 20; Mismatches 117; Indels 84; Gaps 17;
 QY 1 QGGGG--GQGGGGGQGGG--GGGFTALASLASSFMNSNNNNQGGG----- 43
 DB 3033 QGIGSTGQNVPGCYQGQGGGQGGSTPQ--AQQQQPNQYSSNQVGGQAQGGGQPLNQ 3091
 QY 44 -NQSAGSSFGALASSMSSFMNSNNNNNSNQGYNSQYNGNONSQGYNNQGYGGNG 102
 DB 3092 VSGSGSAGAGR-PSQDAVQSGYNGTG-----NQSYRPDQGGQGGQNGQMSGNG 3143
 QY 103 GYQY-----QGGSGCAFFSLAS-----MAOS--YLG-GGCTOSNQGYNDQGG 143
 DB 3144 AQNLRSQAQAQPLQPPQSQPQPAQAQAKNPMASQAYGCGGQGGQYDQ--QGGQ 3201
 QY 144 NNQGYQY--QGGQNYHQGGQGGGQGGSSFSFALASSMSS-----YLG 186
 DB 3202 IAPQAQNPQASQY--GQQQTOQNRYGKSSGTTANSQSSNILLNMSSESGLNQGFSGA 3260
 QY 187 NSNSNSYSYG--QQANQYGRPHNQGSNEYGRPOYG-----GNONSQGHES--- 234
 DB 3261 SSVASSQGGSSQMGSGYGPNGMGQMGQNGQYQRCRPTMGQTNMGQSGMGSGMG 3320
 QY 235 -----FNFSGNFSQNNNGNQ 250
 DB 3321 QTCMSRSLGGLGGIGGGGQSQ 3342

RESULT 3
 ID 019662 PRELIMINARY: PRT: 382 AA.

AC 019662;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
 DE SIMILARITY TO E. COLI CELL DIVISION PROTEIN FTSK.
 GN F2IC10.8.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
 Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 Craxton M., Dear S., Du Z., Durbin R., Favell A., Fulton L.,
 Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 Parsons J., Percy C., Ritken L., Roopra A., Saunders D., Showkhen R.,
 Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,
 Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 Watson A., Wellstock L., Wilkinson-Spoot J., Wohlman P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans.";
 RL Nature 368:32-38(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Du Z., Gatlung S.;
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Waterston R.;
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U55364; AAA97972.1; -
 KW Cell division.
 SO SEQUENCE 382 AA; 42877 MW; 3BD9EB544BBFBF2C CRC64;

Query Match 17.6%; Score 237; DB 5; Length 382;
 Best Local Similarity 28.7%; Pred. No. 2e-13;
 Matches 83; Conservative 32; Mismatches 92; Indels 82; Gaps 13;
 QY 15 QGGGGSFALASLASSFMNSNNNNQGGGSSFGALASSMSSFMNSNNNNNS 74
 DB 73 QNQQYGS-----TAQYNQNMQNMQNSNQYSTTSM-----NNQNSNQY 112
 QY 75 QGGYNSQY-----QNGNONS-----QGYNNQYQGGNGGYQGGGCAFFSLASMA 122
 DB 113 QGQGGNNYGTQADQADQNMQSGQYQGSNNQYNSQNMQMSQSSASG--SSVSILM 169
 QY 123 QSYLGGQGTQNSQOQYNNQGGNNQ--QYQYQGGQNYQH--QGGQGGQGGH-----SSSFS 174
 DB 170 DYSFNNNGNCQYKQDYIENGQTRQATQDELQYQYRNSVDQYMNQYGVGQWINSVFQ 229
 QY 175 ALASSMSSY-----LGNNNSNS-----SYGGQ-----QQ 199
 DB 230 TLPSYVNSFPNIPMTPTAPCLCNAQNGNTTNAIGINTPQARDAYGNQMSNSQFQ 289
 QY 200 ANEYGRPHNG--QQGSNEYGRPHNQGSNEYGRPOYG-----GNONSQGHES--- 247
 DB 290 ANQNGQYQNTQYQOQNGQNGQNF-----QNSNGQNSQYQQTNSQNMNN 335

RESULT 4
 ID 019414 PRELIMINARY: PRT: 409 AA.
 AC 019414;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

Dd	559	SNCGSS-----SNGSSSSNCGSSSSNCGSS---	SNGSSSSNCGSS	600
QY	172	SFSLAAMASSTYLGNNSNSNYGGQQOANETGRPHNQ----	QGSNEYGRPYGGMQ	2280
		: : : : : : : : : : : :		
Dd	601	NOSSSSSNCGSSNCGSSSSNCGSSSSNCGSSSSNCGSSSSNCGSS	6600	
QY	227	NSMGOHSEFNFGNFSGNFSQONNNGN	249	
		: : : : :		
Dd	661	NOSMSGTSSSSNCGSSSSNNN	683	
RESULT	6			
Q9BLX2		PRELIMINARY;	PRT;	929 AA.
AC	Q9BLX2;			
DT	01-JUN-2001 (TREMBLrel. 17, Created)			
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	STATC. PROTEIN.			
GN	DSRC.			
CS	Dictyostelium discoideum (Slime mold).			
CC	Eukaryota; Mycetozoa; Dictyostelida; Dictyostellium.			
CX	NCBI_taxid=44689;			
	[1]			
RN	SEQUENCE FROM N.A.			
FP	STRAIN=CV. AX2;			
FC	Fukuzawa M., Araki T., Adrian I., Williams J.G.;			
FA	"A Dictyostelium SPAT protein that is activated by the stalk cell			
FT	inducer DIF regulates multiple developmental events.";			
EL	Submitted (NOV'2000) to the EMBL/GenBank/DBJ databases.			
EMBL	AJ301670; CAC33514.1; -			
QD	SEQUENCE 929 AA; 106988 MW; ADC150743C1B0233 CRC64;			

Query Match	16.7%	Score 225.5;	DB 5;	Length 929;
Best Local Similarity	26.1%	Pred. No. 4.6e-12;		
Matches 89;	Conservative 23;	Mismatches 100;		
			Indels 129;	Gaps 10

[illegible]

OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi.
 OC Cypriniformes; Cyprinidae; Cyprininae; Cyprinus.
 OX NCBI_TaxID:7962;
 RP [1]
 RP SEQUENCE FROM N.A.
 RA Chang Y.-S., Huang F.-L.;
 RT "Transglutaminase activity is required to recruit fibroblast-like
 RT Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 RL EMBL: AF309414; AAC35716.1; -
 DR InterPro: IPR002221; WAP.
 DR Pfam: PF00095; WAP: 1.
 DR PRINTS: PRO0003; 4DISULPHCORE.
 DR SMART: SM00217; WAP: 1.
 DR PROSITE: PS00317; 4_DISULFIDE_CORE: 1.
 SO SEQUENCE 421 AA; 36510 MW; 999696CF2756CFED7 CRC64;

Query Match	16.6%;	Score 224;	DB 13;	Length 421;
Best Local Similarity	30.4%;	Pred. No. 3e-12;		
Matches	84;	Conservative 12;	Mismatches 144;	Indels 36;
			Gaps	11;

[illegible]

RESULT	8		
043989			
ID	043989	PRELIMINARY:	PRT: 534 AA.
AC	043989;		
DT	01-JUN-1998 (TREMBLrel. 06, Created)		
DT	01-JUN-1998 (TREMBLrel. 06, Last sequence update)		
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)		
DE	HOMEBOX-CONTAINING PROTEIN (FRAGMENT).		
GN	HBX-2.		
OS	Dictyostellium discoideum (SLime mold).		
OC	Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.		
OX	NCBI_TaxId=44689;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=KAX3;		
RA	Han Z., Firtel R.A.;		
RL	Development 0:0-0(1998).		
DR	EMBL, AF036171; AAB92246.1; -		
DR	InterPro; IPR001356; Homeobox.		
DR	Pfam; PF00046; homeobox; 1.		
DR	PROSITE; PS00027; HOMEBOX.1; UNKNOWN.1.		
DR	PROSITE; PS50071; HOMEBOX_2; 1.		
DR	SMART; SM00389; HOX; 1.		
FT	Homeobox; DNA-binding; Nuclear protein.		
FT	NON_TER 534		
SQ	SEQUENCE 534 AA; 63210 MW; 0132653DD65FCFA8 CRC64		

Query Match	16.5%;	Score 222;	DB 5;	Length 534;
Best Local Similarity	23.3%;	Pred. No. 5.6e-12;		

AC 044011;
 DT 01-JUN-1998 (TREMBlrel. 06, Created)
 DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
 DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 GN PROTEIN KINASE YAKA.
 OS Dictyostelium discoideum (Slime mold).
 OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
 ON NCBI_Taxid=44689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-AK800;
 RX MEDLINE=96042901; PubMed=8536963;
 RA Loomis W.F., Welker D., Hughes J., Maghakian D., Kuspa A.;
 RT "Integrated maps of the chromosomes in Dictyostelium discoideum."; Genetics 141:147-157(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-AK800;
 RX MEDLINE=96224325; PubMed=8643615;
 RA Kuspa A., Loomis W.F.;
 RT "Ordered yeast artificial chromosome clones representing the Dictyostelium discoideum genome."; Proc. Natl. Acad. Sci. U.S.A. 93:5562-5566(1996).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-AK800;
 RA Kuspa A., Lu S., Souza G.M.;
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: TO THE SER/THR FAMILY OF PROTEIN KINASES.
 DR EMBL; AF045453; AAC02554.1; -;
 DR HSSP; P24941; 1CKP.
 DR InterPro: IPR000719; Euk.pkinase.
 DR InterPro: IPR002290; Ser_thr_kin_actsite.
 DR Pfam: PF00069; pkinase; 2.
 DR SMART; SM00220; S_TKc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS0108; PROTEIN_KINASE_ST; 1.
 KW ATP-binding; Serine/threonine-protein kinase; Transferase.
 SO SEQUENCE 1457 AA; 167110 MW; C1FCDC99D561856 C64;

Query Match 16.2%; Score 218.5; DB 5; Length 1457;
 Best Local Similarity 29.3%; Pred. No. 2.9e-11;
 Matches 77; Conservative 25; Mismatches 98; Indels 63; Gaps 9;

QY 16 GCGGSGFTALASLASS-----FMNSNNNOGQNGSSGSGFALASMASSFM 63
 DB 789 GSSWGDSSSISLNPSTPTQKQMFQOOQYNNNNNNNNNNNNNNNNNNNN-- 846
 QY 64 HSNNNONSNS-----QCYNSYNGNGNSGYNNOQYQCG 100
 DB 847 NNNNNNNRRNRKSDIPDSFSSSEGMDFNLYYQOQOQOQOQOQOQOQOQ 906
 QY 101 NGCYOQOQGGGAFSSLASM-----AGSYLGAGTQSNQOQYNOQOQON 144
 DB 907 QOQ 964
 QY 145 NQOQYQOQ 203
 DB 965 NTCOQ 1023
 QY 204 GRPO---HNGGQOQSNEX---GRP 220
 DB 1024 KQOQKMLHHQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQ 1046

RESULT 12
 ID 041553 PRELIMINARY; PRT; 815 AA.
 AC 041553;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)

DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE TRICRYCUM AESTIVUM HIGH MW GLUTENIN SUBUNIT (AX2+).
 OS Triticum aestivum (Wheat).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
 OC Triticeae; Triticum.
 ON NCBI_Taxid=4565;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. CHEYENNE;
 RA Anderson O.D., Greene F.C.;
 RL Theor. Appl. Genet. 77:689-700(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. CHEYENNE;
 RA Anderson O.D.;
 RL Submitted (JAN-1989) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. CHEYENNE;
 RA Anderson O.D.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL; M22208; AAB02788.1; -;
 DR Mendel; 14743; Trlae;2380;14743.
 DR InterPro: IPR001419; Glutennin.
 DR PRINTS; PRO02014; GLUTENIN.
 SO SEQUENCE 815 AA; 88476 MW; FC09905398EA71A8 CRC64;

Query Match 16.2%; Score 218; DB 10; Length 815;
 Best Local Similarity 31.9%; Pred. No. 1.9e-11;
 Matches 87; Conservative 13; Mismatches 125; Indels 48; Gaps 12;

QY 2 GCGGCGGCGG-QCGGCGGCGGFTALASLASSFMNSNNNOGQNGSSGSGFALASMAS 60
 DB 198 GCGGQLQCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 248
 QY 61 SFMNSNNONSNSQCGY--NOSYONGNONSQGYN-----NQYGGNG--GY----- 104
 DB 249 --QOSGCGQLGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 305
 QY 105 -QOQOQGGGAFSSLASMASSYLGGGQTOQSNQOQYNOQ-GONNOQYQOQOQOQOQOQ 162
 DB 306 QOSGCGGSGGYRPSAQQPGQLQSTQEQQLGQEQDDQDSGCG--RQCGGCGGQRRQDDQSG 363
 QY 163 QOQOQGHSSSFALASMASSYLGNNSNSSSYGGGQOQANEXGRPOHNG-QOSNEXGRPO 221
 DB 364 QGQDPGGRP-----GYSTSPQQLGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 411
 QY 222 YG--GNQNSNGCHSFNFGNFGSQNNNGNQR 252
 DB 412 QGQGQGPQEGQCGGQCGGQCGGQCGGQCGGQCGGQCGGQCGGQCGGQCGGQCGG 444

RESULT 13
 ID 003872 PRELIMINARY; PRT; 830 AA.
 AC 003872;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE HIGH MOLECULAR WEIGHT GLUTENIN SUBUNIT 1A1 PRECURSOR.
 GN GLU-1A1.
 OS Triticum aestivum (Wheat).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
 OC Triticeae; Triticum.
 ON NCBI_Taxid=4565;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. HOPE; TISSUE=ENDOSPERM;
 RA Halford N.G., Field J.M., Blair H., Urwin P., Moore K., Robert L.,


```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See license@isb-sib.ch).
CC -----
DR EMBL: Z11327; CAA77476.1; -.
DR PIR: S18043; S18043.
DR PIR: S20907; S20907.
DR HSSP: P48793; 1XND.
DR InterPro: IPR001000; Glyco_hydro_10.
DR InterPro: IPR01137; Glyco_hydro_11.
DR Pfam: PF00331; Glyco_hydro_10_1.
DR Pfam: PF00457; Glyco_hydro_11_1.
DR PRINTS: PR00134; GLHYDRASE10.
DR PRINTS: PR00091; GLHYDRASE11.
DR PROSITE: PS00591; GLYCOSYL_HYDROL_F10_1.
DR PROSITE: PS00776; GLYCOSYL_HYDROL_F11_1.
DR PROSITE: PS00777; GLYCOSYL_HYDROL_F11_2; 1.
DR Xylan degradation; Hydrolase; Glycosidase; Multifunctional enzyme;
KW Repeat; signal.
FT SIGNAL 1 27
FT CHAIN 28 954
FT DOMAIN 28 244
FT DOMAIN 245 622
FT DOMAIN 623 954
FT ACT_SITE 122 122
FT ACT_SITE 223 223
FT ACT_SITE 774 774
FT ACT_SITE 884 884
FT ACT_SITE 954 AA; 11362 MW; 1033567D4B526EBD CRC64;
SO SEQUENCE

Query Match 16.2%; Score 218; DB 1; Length 954;
Best Local Similarity 26.2%; Pred. No. 3.5e-07;
Matches 80; Conservative 44; Mismatches 95; Indels 86; Gaps 15;

QY 1 QGQGQ-----GQ-----GQGQGQGQGSGFTALASLASSFMSNNN-----NOGQND 46
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 307 QGQGGNNNDMMNMGQNDNDMMNMGQGGQNDNDM-----NMNGQNDNDMMNMGQGGQND 360
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 47 SGGSSFGALASSFMSNNNNSNSQGYN--QSTYONGNSQYNNQYQG--GNGG 103
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 361 NDMMNMG-----QNDNDMMNMGNN--NQANWNGDNDNNNNNQNDQNDNDMMNMGNN 415
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 104 YQGGQGGGGAFFSLASSMAQSTLGGGQYQNSQ--QYNNQGGQNNQYQYQGGQNTYHQDQG 162
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 416 NNQGGQNDNDMMNMGNN-----QNNNNNNQNDNDMMNMGQNDNDMMNMGNNNNNN 470
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 163 QQ-----QQQGSSSFGALASSMASSYLGNNS---NSNSYQG 196
    || || || || || || || || || || || || || || || || || || || || ||
DB 471 QQMANNGDNNNNNNMMQNDNNQNDMMNMGQ-----NNTWDMNNQNNNN 512
    || || || || || || || || || || || || || || || || || || || || ||
QY 197 QQQANEYRPOHNGQQGQSNEXGRPOYGQNSNG--QHESEFNSGFNQD-----NNG 248
    || : : : : : ||||| : : : : : || : : : : : || : : : : : || : : : : :
DB 513 NQNNNDMMNMGNNQGGQNDNDMMNMGNNNNQNNQNDMMNMGNNQNDMMNMGNNQNDMMN 572
    || : : : : : ||||| : : : : : || : : : : : || : : : : : || : : : : :
QY 249 NQNR 253
    || : : :
DB 573 NNQNW 577
    || : : :

RESULT 4
GLTS_WHEAT STANDARD: PRT: 839 AA.
AC P10388;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE GLUTENIN, HIGH MOLECULAR WEIGHT SUBUNIT DX5 PRECURSOR.
OS GLU-ID-1D OR GLU-D1-1B.
   Triticum aestivum (wheat).

```

	OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
	CC	Spermatophytas; Magnoliophyta; Liliopsida; Poales; Rosaceae; Pooideae;
	OC	Triticeae; Triticum.
	CX	NCBL MaxID=4565;
RN	(1)	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=cv, CHEYENNE:	
RX	MEDLINE=69089419; PubMed=2563152;	
RA	Andereson O.D., Greene F.C., Yip R.E., Halford N.G., Shewry P.R., Malpica-Romero J.M.;	
RT	"Nucleotide sequences of the two high-molecular-weight glutenin genes from the D-genome of a hexaploid bread wheat, Triticum aestivum L. cv Cheyenne";	
RL	Nucleic Acids Res. 17:461-462(1989).	
RN	[2]	
RP	REVISIONS, SEQUENCE FROM N.A.	
RC	STRAIN=cv, CHEYENNE:	
RA	Andereson O.D. ;	
RL	Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.	
CC	-I- FUNCTION: GLUTEININS ARE THE HIGH MOLECULAR WEIGHT SED STORAGE PROTEIN OF WHEAT ENDOSPERM. THOUGHT TO BE RESPONSIBLE FOR THE VISCO-ELASTIC PROPERTY OF WHEAT DOUGH. -I- SUBUNIT: DISULFIDE-BRIDGE LINKED AGGREGATES. -I- MISCELLANEOUS: GLUTENINS ARE CODED BY SEVERAL GENES ON EACH OF THE GROUP 1 CHROMOSOMES OF WHEAT. -I- MISCELLANEOUS: THE MATURE PROTEIN IS CHARACTERIZED BY A LARGE NUMBER OF WELL PRESERVED REPEATS OF THE TWO MOTIFS: GQDPGO AND GOOPGGOGGYPTS. ----- -----	
CC	This SWISS-Prot entry is copyright. It is produced through a collaboration between The Swiss Institute of Bioinformatics, There are no restrictions on the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-slb.ch/announce/ or send an email to licensel@isb-slb.ch). ----- -----	
DR	EMBL; X12928; CAA31395.1; -.	
DR	PIR; S02262; S02262.	
DR	InterPro; IPR001419; GlutenIn.	
DR	PRINTS; PR00210; GLUTENIN.	
KW	Seed storage protein; Repeat; Multigene family; Signal.	
FT SIGNAL	1 21	GLUTENIN, HIGH MOLECULAR WEIGHT SUBUNIT
FT FT	22 839	Dx5.
FT FT		REPEATS.
SO DOMAIN	131 801	DX5.
SEQUENCE	839 AA; 89359 MW; 0FL4EI106DD552643 CRC64;	
	Query Match	16.1%; Score 217.5; DB 1; Length 839;
	Best Local Similarity	29.9%; Pred. No. 3.4e-07;
	Matches	90; Conservative 13; Mismatches 133; Indels 65; Gaps 11;
OY	1 QCGGCGCQGCGGCGGCGGTALASLSFMNSNNNQGNQSGSSFGALSAMAS 60	
		:: :: :
Db	184 OGQQGGCG-QGQGGCGG-----PGYPPTSSQLDFPOLLQPAGQGGQPPGAQ 232	
OY	61 SFMHSHNNNSSNGSYNGSYONGNONMSGYNNOYYGGNGCY-----QQNGGSOSA 114	
		:
Db	233 GGQRPGGQQPQGQGQG--QDPGGQQQPDQGGGGQLDGQGQGYPTSLDSGGGSPGY 290	
OY	115 FSLSMASMSYLK-----GCOTQSNOQDYNOQGO--NNQDDYQDQGVYQHQQGQ 163	
	: : : :	:
Db	291 PTSLDQDLQGSGGSGYPTSPQGPGRGQGGQLQPAQGGQPPGGGQGGQGDQDPGQ 350	
OY	164 QQQGGHSSFSAIASMASSYLK-----NNSNSNSTYGQ-----QANEYGR 205	
		:
Db	351 GQQPFQGGPYTPPSQSGSGGQPGYPPTSSSQDPQSSQPCGGGQGGVGQDAQAPFG 410	
OY	206 -----POINGQDSMEY-GRPDYGGNMNSNGCHSFNFSGNSEQNNGCN 249	
	: :	:
Db	411 GQQPEGQGPgyTPTSPQSGGQGGPEgYTLTSPQSSSQGQPPGCLAQ-SAQGAKGGQPPGGQ 469	

Db	228	OPOOOHHHNOOQOHNOHQOHOQQOONQIOMVPOOPQOSLNSGNNNNNNNNNNNNNNNNNNN	287
Qy	249	NNN 251	
Db	288	NNN 290	
<hr/>			
RESULT 10			
FUS_MOUSE	ID	FUS_MOUSE STANDARD; PRT; 518 AA.	
AC	P56959;		
Dt	20-AUG-2001	(Rel. 40, Created)	
Dt	20-AUG-2001	(Rel. 40, Last sequence update)	
Dt	20-AUG-2001	(Rel. 40, Last annotation update)	
DE	RNA-BINDING PROTEIN FUS (PIGEPEN PROTEIN).		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OX	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
NCBI_Taxid=10090;			
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Ailaprat S.R., Zhang M., Zhao X., Alliegro M.A., Alliegro M.C.,		
RA	Burdal C.A.;		
RL	"Regulation of pligen expression in mouse embryos."		
CC	Submitted (JAN-2000) to the EMBL/Genbank/DBU databases.		
CC	-1- FUNCTION: BINDS BOTH SINGLE-STRADED AND DOUBLE-STRADED DNA AND		
CC	PROMOTES ATP-INDEPENDENT ANNEALING OF COMPLEMENTARY SINGLE-		
CC	STRANDED DNAs AND D-LOOP FORMATION IN SUPERHELICAL DOUBLE-STRADED		
CC	DNA. MAY PLAY A ROLE IN MAINTENANCE OF GENOMIC INTEGRITY (BY		
CC	SIMILARITY).		
CC	-1- SUBUNIT: COMPONENT OF NUCLEAR RIBOPROTEIN COMPLEXES (BY		
CC	SIMILARITY).		
CC	-1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).		
CC	-1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).		
CC	-1- SIMILARITY: BELONGS TO THE TEN FAMILY OF RNP PROTEINS.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; AF224264; AAF70602.1; -		
DR	MGI; MG1:1353633; FUS.		
DR	InterPro; IPR002952; Eggshell.		
DR	InterPro; IPR005054; RRM.		
DR	InterPro; IPR001876; Znf-RanBP.		
DR	Pfam; PF00076; Trm; 1.		
DR	Pfam; PF00641; zf-RanBP; 1.		
DR	SMART; SM00360; RRM; 1.		
DR	SMART; SM00547; ZnF_RBZ; 1.		
DR	PROSITE; PS50102; RRM; 1.		
DR	PROSITE; PS00030; RRM_NRP_1; FALSE_NEG.		
KW	RNA-binding; DNA-binding; Nuclear protein; Repeat; Zinc-finger; Zinc;		
KW	Metal-binding.		
FT	DOMAIN 1 167 GLN/GLY/SER/TYR-RICH.		
FT	DOMAIN 168 265 GLY-RICH.		
FT	DOMAIN 278 364 RNA-BINDING (RRM).		
FT	DOMAIN 364 518 ARG/GLY-RICH.		
FT	ZN-FING 421 440 C4-TYPE (POTENTIAL).		
SO	SEQUENCE 518 AA; 52673 MW; E06FE231BREED78D6 CNC64;		
<hr/>			
Query Match	15.0%;	Score 202;	DB 1; Length 518;
Best Local Similarity	30.1%;	Pred. No. 2e-06;	
Matches	77;	Conservative 25;	Mismatches 120; Indels 34; Gaps 10;
Yy	6	GGGGGGGGGGGGGGGGCTATLASLASSFFNNNNNNNNNNNNNNNNNNNNNGSSGSFFALASMASSFFMHS	65
Yy			

Db	22	GAGGVSQSSQSPBYGGQSSYGSGGASDT-----SGYGSSSYGSSSYG-----	OTQ	63
Qy	66	NNNNNNNSQGGYNSYONG--NNSQGYNNQYQSGNGNGYQQQQQQSGGAFSSLASMAQS	124	
Db	64	NTGYGTQAPAGYGTGGTGGYSSQSSSYSGYGGQPPAPBSSTGGYCG--SSQSS	122	
Qy	125	YLGGGQGTQSSNQQQYMQQ--GQNNMQQ--YQQGQGMVYQHQQGQGGQ0000GHSSFSFALASMASS	182	
Db	123	Y---GQPGSG--GYQQQSGYGGGQQQSYQQQSSSYPPPGYGGQNNYNNSSSGGGGGGGGN	177	
Qy	183	YLGNNNSNS-----SYGQGGQQANEYRPHQNGQGSNEGPRQYGGNNSNGCHESFN	236	
Db	178	Y-GQQQSSMSGGGGGGGYGGQNDQSSGGGGGYGGGQDRCGRGRGGGGYNRSSGGYEPRG	236	
Qy	237	FSGNFSQQNNNGNQR	252	
Db	237	RGGGRGGRGGMGGSDR	252	
RESULT 11				
FUS_BOVIN	1D	FUS_BOVIN	STANDARD;	PRT; 512 AA.
AC	Q28009;	01-NOV-1997 (Rel. 35, Created)		
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	RNA-BINDING PROTEIN FUS (PIGPEN PROTEIN).			
GN	FUS.			
OS	Bos taurus (Bovine).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
CC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;			
OC	Bovidae; Bovinae; Bos.			
OX	NCBI_Taxid:9913;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE-Aorta;			
RX	MEDLINE=96175600; PubMed=8631501;			
RA	Alliegro M.C., Alliegro M.A.;			
RT	"A nuclear protein regulated during the transition from active to			
RL	quiescent phenotype in cultured endothelial cells.";			
RL	Dev. Biol. 174:288-297(1996).			
RP	[2]			
RC	CARBOHYDRATE BINDING DOMAIN			
RX	MEDLINE=20160719; PubMed=10694442;			
RA	Alliegro M.C.;			
RT	"A C-terminal carbohydrate-binding domain in the endothelial cell			
RL	regulatory protein, Pigpen: new function for an EWS family member";			
RL	Regul. Cell Res. 255:270-277(2000).			
CC	-1- FUNCTION: BINDS BOTH SINGLE-STRANDED AND DOUBLE-STRANDED DNA AND			
CC	PROMOTES ATP-INDEPENDENT ANNEALING OF COMPLEMENTARY SINGLE-			
CC	STRANDED DNAs AND D-LOOP FORMATION IN SUPERHELICAL DOUBLE-STRANDED			
CC	DNA. MAY PLAY A ROLE IN MAINTENANCE OF GENOMIC INTEGRITY (BY			
CC	SIMILARITY).			
CC	-1- SUBUNIT: COMPONENT OF NUCLEAR RIBOPROTEIN COMPLEXES (BY			
CC	SIMILARITY).			
CC	-1- SUBCELLULAR LOCATION: NUCLEAR. EXHIBITS DIFFUSE STAINING			
CC	THROUGHOUT (EXCLUDING NUCLEOLI), TOGETHER WITH A SMALL NUMBER OF			
CC	INTENSELY STAINED FOCAL POINTS, OR GRANULES, AND PUNCTATE STAINING			
CC	ALONG THE NUCLEAR ENVELOPE.			
CC	-1- DOMAIN: THE C-TERMINAL DOMAIN BINDS CARBOHYDRATES.			
CC	-1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).			
CC	-1- SIMILARITY: BELONGS TO THE TET FAMILY OF RNP PROTEINS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation			
CC	the European Bioinformatics Institute. There are no restrictions on			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed, usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@sib-sib.ch).			
CC	-----			
CC	EMBL; U26024; AAC13543.1; -.			
DR	HSSP; P09651.1HA1.			

QY 207 OHNGOQSNEXGRPOYGNONSNGHSEFNFSGNFQNNNGNQ 250
 DB 479 ---GORGQPGGQHPEGGQPGGGQGGYPTG---PQGPGGGQ 516

RESULT 15
 RB56_HUMAN
 ID RB56_HUMAN STANDARD; PRT; 592 AA.
 AC Q92804; Q92871;
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE TATA-BINDING PROTEIN ASSOCIATED FACTOR 2N (RNA-BINDING PROTEIN 56)
 DE (TAFII168) (TAF(II)68).
 GN TAF2N OR RBP56.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS).
 RX MEDLINE=97124846; PubMed=8954779;
 RA Morohoshi F., Aral K., Takahashi E.-I., Tanigami A., Ohki M.;
 RT "Cloning and mapping of a human RBP56 gene encoding a putative RNA
 RT binding protein similar to FUS/TLS and EWS proteins.";
 RL Genomics 38:51-57(1996).
 RN [2]
 RP SEQUENCE FROM N.A. (SHORT ISOFORM), AND SEQUENCE OF 282-297 & 307-320.
 RX MEDLINE=97045110; PubMed=8890175;
 RA Bertolotti A., Lutz Y., Heard D.J., Chambon P., Tora L.;
 RT "hTAF(II)68, a novel RNA/ssDNA-binding protein with homology to the
 RT pro-oncoteleins TLS/FUS and EWS is associated with both TFIID and RNA
 RT polymerase II.";
 RL EMBO J. 15:5022-5031(1996).
 RN [3]
 RP SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS).
 RX MEDLINE=99013873; PubMed=9795213;
 RA Morohoshi F., Ootsuka Y., Aral K., Ichikawa H., Mitani S.,
 RA Munakata M., Ohki M.;
 RT "Genomic structure of the human RBP56/hTAFII68 and FUS/TLS genes.";
 RL Gene 221:191-198(1998).
 CC -1- FUNCTION: RNA AND ssDNA-BINDING PROTEIN THAT MAY PLAY SPECIFIC
 CC ROLES DURING TRANSCRIPTION INITIATION AT DISTINCT PROMOTERS. CAN
 CC ENTER THE PREINITIATION COMPLEX TOGETHER WITH THE RNA POLYMERASE
 CC II (POL. II).
 CC -1- SUBUNIT: BELONGS TO THE RNA POLYMERASE II (POL. II) TRANSCRIPTIONAL
 CC MULTIPROTEIN COMPLEX, TOGETHER WITH THE TATA-BINDING PROTEIN (TBP)
 CC AND OTHER TBP-ASSOCIATED FACTORS (TAF(II)S).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG ISOFORM (SHOWN HERE)
 CC AND A SHORT ISOFORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: UBIQUITOUS. OBSERVED IN ALL FETAL AND ADULT
 CC TISSUES.
 CC -1- DISEASE: A FORM OF EXTRASKELETAL MYXOID CHONDROSARCOMA (EMC) IS
 CC CHARACTERIZED BY A CHROMOSOMAL TRANSLOCATION T(9;17)(Q22;Q11) THAT
 CC INVOLVES TAF2N AND NR4A3.
 CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RNM).
 CC -1- SIMILARITY: CONTAINS 1 RANBP-TYPE ZINC FINGER.
 CC -1- SIMILARITY: BELONGS TO THE TET FAMILY OF RNP PROTEINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-db.ch).
 CC
 CC EMBL: U51334; AAC50932.1; -
 CC EMBL: X98893; CAA67398.1; -
 CC EMBL: AB010067; BAA33811.1; -
 CC EMBL: AB010057; BAA33811.1; JOINED.

DR EMBL: AB010058; BAA33811.1; JOINED.
 DR EMBL: AB010059; BAA33811.1; JOINED.
 DR EMBL: AB010060; BAA33811.1; JOINED.
 DR EMBL: AB010061; BAA33811.1; JOINED.
 DR EMBL: AB010062; BAA33811.1; JOINED.
 DR EMBL: AB010063; BAA33811.1; JOINED.
 DR EMBL: AB010064; BAA33811.1; JOINED.
 DR EMBL: AB010065; BAA33811.1; JOINED.
 DR EMBL: AB010066; BAA33811.1; JOINED.
 DR EMBL: AB010067; BAA33811.1; JOINED.
 DR EMBL: AB010068; BAA33812.1; JOINED.
 DR EMBL: AB010069; BAA33812.1; JOINED.
 DR EMBL: AB010070; BAA33812.1; JOINED.
 DR EMBL: AB010071; BAA33812.1; JOINED.
 DR EMBL: AB010072; BAA33812.1; JOINED.
 DR EMBL: AB010073; BAA33812.1; JOINED.
 DR EMBL: AB010074; BAA33812.1; JOINED.
 DR EMBL: AB010075; BAA33812.1; JOINED.
 DR EMBL: AB010076; BAA33812.1; JOINED.
 DR EMBL: AB010077; BAA33812.1; JOINED.
 DR EMBL: AB010078; BAA33812.1; JOINED.
 DR EMBL: AB010079; BAA33812.1; JOINED.
 DR EMBL: AB010080; BAA33812.1; JOINED.
 DR EMBL: AB010081; BAA33812.1; JOINED.
 DR EMBL: AB010082; BAA33812.1; JOINED.
 DR EMBL: AB010083; BAA33812.1; JOINED.
 DR EMBL: AB010084; BAA33812.1; JOINED.
 DR EMBL: AB010085; BAA33812.1; JOINED.
 DR EMBL: AB010086; BAA33812.1; JOINED.
 DR EMBL: AB010087; BAA33812.1; JOINED.
 DR EMBL: AB010088; BAA33812.1; JOINED.
 DR EMBL: AB010089; BAA33812.1; JOINED.
 DR EMBL: AB010090; BAA33812.1; JOINED.
 DR EMBL: AB010091; BAA33812.1; JOINED.
 DR EMBL: AB010092; BAA33812.1; JOINED.
 DR EMBL: AB010093; BAA33812.1; JOINED.
 DR EMBL: AB010094; BAA33812.1; JOINED.
 DR EMBL: AB010095; BAA33812.1; JOINED.
 DR EMBL: AB010096; BAA33812.1; JOINED.
 DR EMBL: AB010097; BAA33812.1; JOINED.
 DR EMBL: AB010098; BAA33812.1; JOINED.
 DR EMBL: AB010099; BAA33812.1; JOINED.
 DR EMBL: AB010100; BAA33812.1; JOINED.
 DR EMBL: AB010101; BAA33812.1; JOINED.
 DR EMBL: AB010102; BAA33812.1; JOINED.
 DR EMBL: AB010103; BAA33812.1; JOINED.
 DR EMBL: AB010104; BAA33812.1; JOINED.
 DR EMBL: AB010105; BAA33812.1; JOINED.
 DR EMBL: AB010106; BAA33812.1; JOINED.
 DR EMBL: AB010107; BAA33812.1; JOINED.
 DR EMBL: AB010108; BAA33812.1; JOINED.
 DR EMBL: AB010109; BAA33812.1; JOINED.
 DR EMBL: AB010110; BAA33812.1; JOINED.
 DR EMBL: AB010111; BAA33812.1; JOINED.
 DR EMBL: AB010112; BAA33812.1; JOINED.
 DR EMBL: AB010113; BAA33812.1; JOINED.
 DR EMBL: AB010114; BAA33812.1; JOINED.
 DR EMBL: AB010115; BAA33812.1; JOINED.
 DR EMBL: AB010116; BAA33812.1; JOINED.
 DR EMBL: AB010117; BAA33812.1; JOINED.
 DR EMBL: AB010118; BAA33812.1; JOINED.
 DR EMBL: AB010119; BAA33812.1; JOINED.
 DR EMBL: AB010120; BAA33812.1; JOINED.
 DR EMBL: AB010121; BAA33812.1; JOINED.
 DR EMBL: AB010122; BAA33812.1; JOINED.
 DR EMBL: AB010123; BAA33812.1; JOINED.
 DR EMBL: AB010124; BAA33812.1; JOINED.
 DR EMBL: AB010125; BAA33812.1; JOINED.
 DR EMBL: AB010126; BAA33812.1; JOINED.
 DR EMBL: AB010127; BAA33812.1; JOINED.
 DR EMBL: AB010128; BAA33812.1; JOINED.
 DR EMBL: AB010129; BAA33812.1; JOINED.
 DR EMBL: AB010130; BAA33812.1; JOINED.
 DR EMBL: AB010131; BAA33812.1; JOINED.
 DR EMBL: AB010132; BAA33812.1; JOINED.
 DR EMBL: AB010133; BAA33812.1; JOINED.
 DR EMBL: AB010134; BAA33812.1; JOINED.
 DR EMBL: AB010135; BAA33812.1; JOINED.
 DR EMBL: AB010136; BAA33812.1; JOINED.
 DR EMBL: AB010137; BAA33812.1; JOINED.
 DR EMBL: AB010138; BAA33812.1; JOINED.
 DR EMBL: AB010139; BAA33812.1; JOINED.
 DR EMBL: AB010140; BAA33812.1; JOINED.
 DR EMBL: AB010141; BAA33812.1; JOINED.
 DR EMBL: AB010142; BAA33812.1; JOINED.
 DR EMBL: AB010143; BAA33812.1; JOINED.
 DR EMBL: AB010144; BAA33812.1; JOINED.
 DR EMBL: AB010145; BAA33812.1; JOINED.
 DR EMBL: AB010146; BAA33812.1; JOINED.
 DR EMBL: AB010147; BAA33812.1; JOINED.
 DR EMBL: AB010148; BAA33812.1; JOINED.
 DR EMBL: AB010149; BAA33812.1; JOINED.
 DR EMBL: AB010150; BAA33812.1; JOINED.
 DR EMBL: AB010151; BAA33812.1; JOINED.
 DR EMBL: AB010152; BAA33812.1; JOINED.
 DR EMBL: AB010153; BAA33812.1; JOINED.
 DR EMBL: AB010154; BAA33812.1; JOINED.
 DR EMBL: AB010155; BAA33812.1; JOINED.
 DR EMBL: AB010156; BAA33812.1; JOINED.
 DR EMBL: AB010157; BAA33812.1; JOINED.
 DR EMBL: AB010158; BAA33812.1; JOINED.
 DR EMBL: AB010159; BAA33812.1; JOINED.
 DR EMBL: AB010160; BAA33812.1; JOINED.
 DR EMBL: AB010161; BAA33812.1; JOINED.
 DR EMBL: AB010162; BAA33812.1; JOINED.
 DR EMBL: AB010163; BAA33812.1; JOINED.
 DR EMBL: AB010164; BAA33812.1; JOINED.
 DR EMBL: AB010165; BAA33812.1; JOINED.
 DR EMBL: AB010166; BAA33812.1; JOINED.
 DR EMBL: AB010167; BAA33812.1; JOINED.
 DR EMBL: AB010168; BAA33812.1; JOINED.
 DR EMBL: AB010169; BAA33812.1; JOINED.
 DR EMBL: AB010170; BAA33812.1; JOINED.
 DR EMBL: AB010171; BAA33812.1; JOINED.
 DR EMBL: AB010172; BAA33812.1; JOINED.
 DR EMBL: AB010173; BAA33812.1; JOINED.
 DR EMBL: AB010174; BAA33812.1; JOINED.
 DR EMBL: AB010175; BAA33812.1; JOINED.
 DR EMBL: AB010176; BAA33812.1; JOINED.
 DR EMBL: AB010177; BAA33812.1; JOINED.
 DR EMBL: AB010178; BAA33812.1; JOINED.
 DR EMBL: AB010179; BAA33812.1; JOINED.
 DR EMBL: AB010180; BAA33812.1; JOINED.
 DR EMBL: AB010181; BAA33812.1; JOINED.
 DR EMBL: AB010182; BAA33812.1; JOINED.
 DR EMBL: AB010183; BAA33812.1; JOINED.
 DR EMBL: AB010184; BAA33812.1; JOINED.
 DR EMBL: AB010185; BAA33812.1; JOINED.
 DR EMBL: AB010186; BAA33812.1; JOINED.
 DR EMBL: AB010187; BAA33812.1; JOINED.
 DR EMBL: AB010188; BAA33812.1; JOINED.
 DR EMBL: AB010189; BAA33812.1; JOINED.
 DR EMBL: AB010190; BAA33812.1; JOINED.
 DR EMBL: AB010191; BAA33812.1; JOINED.
 DR EMBL: AB010192; BAA33812.1; JOINED.
 DR EMBL: AB010193; BAA33812.1; JOINED.
 DR EMBL: AB010194; BAA33812.1; JOINED.
 DR EMBL: AB010195; BAA33812.1; JOINED.
 DR EMBL: AB010196; BAA33812.1; JOINED.
 DR EMBL: AB010197; BAA33812.1; JOINED.
 DR EMBL: AB010198; BAA33812.1; JOINED.
 DR EMBL: AB010199; BAA33812.1; JOINED.
 DR EMBL: AB010200; BAA33812.1; JOINED.
 DR EMBL: AB010201; BAA33812.1; JOINED.
 DR EMBL: AB010202; BAA33812.1; JOINED.
 DR EMBL: AB010203; BAA33812.1; JOINED.
 DR EMBL: AB010204; BAA33812.1; JOINED.
 DR EMBL: AB010205; BAA33812.1; JOINED.
 DR EMBL: AB010206; BAA33812.1; JOINED.
 DR EMBL: AB010207; BAA33812.1; JOINED.
 DR EMBL: AB010208; BAA33812.1; JOINED.
 DR EMBL: AB010209; BAA33812.1; JOINED.
 DR EMBL: AB010210; BAA33812.1; JOINED.
 DR EMBL: AB010211; BAA33812.1; JOINED.
 DR EMBL: AB010212; BAA33812.1; JOINED.
 DR EMBL: AB010213; BAA33812.1; JOINED.
 DR EMBL: AB010214; BAA33812.1; JOINED.
 DR EMBL: AB010215; BAA33812.1; JOINED.
 DR EMBL: AB010216; BAA33812.1; JOINED.
 DR EMBL: AB010217; BAA33812.1; JOINED.
 DR EMBL: AB010218; BAA33812.1; JOINED.
 DR EMBL: AB010219; BAA33812.1; JOINED.
 DR EMBL: AB010220; BAA33812.1; JOINED.
 DR EMBL: AB010221; BAA33812.1; JOINED.
 DR EMBL: AB010222; BAA33812.1; JOINED.
 DR EMBL: AB010223; BAA33812.1; JOINED.
 DR EMBL: AB010224; BAA33812.1; JOINED.
 DR EMBL: AB010225; BAA33812.1; JOINED.
 DR EMBL: AB010226; BAA33812.1; JOINED.
 DR EMBL: AB010227; BAA33812.1; JOINED.
 DR EMBL: AB010228; BAA33812.1; JOINED.
 DR EMBL: AB010229; BAA33812.1; JOINED.
 DR EMBL: AB010230; BAA33812.1; JOINED.
 DR EMBL: AB010231; BAA33812.1; JOINED.
 DR EMBL: AB010232; BAA33812.1; JOINED.
 DR EMBL: AB010233; BAA33812.1; JOINED.
 DR EMBL: AB010234; BAA33812.1; JOINED.
 DR EMBL: AB010235; BAA33812.1; JOINED.
 DR EMBL: AB010236; BAA33812.1; JOINED.
 DR EMBL: AB010237; BAA33812.1; JOINED.
 DR EMBL: AB010238; BAA33812.1; JOINED.
 DR EMBL: AB010239; BAA33812.1; JOINED.
 DR EMBL: AB010240; BAA33812.1; JOINED.
 DR EMBL: AB010241; BAA33812.1; JOINED.
 DR EMBL: AB010242; BAA33812.1; JOINED.
 DR EMBL: AB010243; BAA33812.1; JOINED.
 DR EMBL: AB010244; BAA33812.1; JOINED.
 DR EMBL: AB010245; BAA33812.1; JOINED.
 DR EMBL: AB010246; BAA33812.1; JOINED.
 DR EMBL: AB010247; BAA33812.1; JOINED.
 DR EMBL: AB010248; BAA33812.1; JOINED.
 DR EMBL: AB010249; BAA33812.1; JOINED.
 DR EMBL: AB010250; BAA33812.1; JOINED.
 DR EMBL: AB010251; BAA33812.1; JOINED.
 DR EMBL: AB010252; BAA33812.1; JOINED.
 DR EMBL: AB010253; BAA33812.1; JOINED.
 DR EMBL: AB010254; BAA33812.1; JOINED.
 DR EMBL: AB010255; BAA33812.1; JOINED.
 DR EMBL: AB010256; BAA33812.1; JOINED.
 DR EMBL: AB010257; BAA33812.1; JOINED.
 DR EMBL: AB010258; BAA33812.1; JOINED.
 DR EMBL: AB010259; BAA33812.1; JOINED.
 DR EMBL: AB010260; BAA33812.1; JOINED.
 DR EMBL: AB010261; BAA33812.1; JOINED.
 DR EMBL: AB010262; BAA33812.1; JOINED.
 DR EMBL: AB010263; BAA33812.1; JOINED.
 DR EMBL: AB010264; BAA33812.1; JOINED.
 DR EMBL: AB010265; BAA33812.1; JOINED.
 DR EMBL: AB010266; BAA33812.1; JOINED.
 DR EMBL: AB010267; BAA33812.1; JOINED.
 DR EMBL: AB010268; BAA33812.1; JOINED.
 DR EMBL: AB010269; BAA33812.1; JOINED.
 DR EMBL: AB010270; BAA33812.1; JOINED.
 DR EMBL: AB010271; BAA33812.1; JOINED.
 DR EMBL: AB010272; BAA33812.1; JOINED.
 DR EMBL: AB010273; BAA33812.1; JOINED.
 DR EMBL: AB010274; BAA33812.1; JOINED.
 DR EMBL: AB010275; BAA33812.1; JOINED.
 DR EMBL: AB010276; BAA33812.1; JOINED.
 DR EMBL: AB010277; BAA33812.1; JOINED.
 DR EMBL: AB010278; BAA33812.1; JOINED.
 DR EMBL: AB010279; BAA33812.1; JOINED.
 DR EMBL: AB010280; BAA33812.1; JOINED.
 DR EMBL: AB010281; BAA33812.1; JOINED.
 DR EMBL: AB010282; BAA33812.1; JOINED.
 DR EMBL: AB010283; BAA33812.1; JOINED.
 DR EMBL: AB010284; BAA33812.1; JOINED.
 DR EMBL: AB010285; BAA33812.1; JOINED.
 DR EMBL: AB010286; BAA33812.1; JOINED.
 DR EMBL: AB010287; BAA33812.1; JOINED.
 DR EMBL: AB010288; BAA33812.1; JOINED.
 DR EMBL: AB010289; BAA33812.1; JOINED.
 DR EMBL: AB010290; BAA33812.1; JOINED.
 DR EMBL: AB010291; BAA33812.1; JOINED.
 DR EMBL: AB010292; BAA33812.1; JOINED.
 DR EMBL: AB010293; BAA33812.1; JOINED.
 DR EMBL: AB010294; BAA33812.1; JOINED.
 DR EMBL: AB010295; BAA33812.1; JOINED.
 DR EMBL: AB010296; BAA33812.1; JOINED.
 DR EMBL: AB010297; BAA33812.1; JOINED.
 DR EMBL: AB010298; BAA33812.1; JOINED.
 DR EMBL: AB010299; BAA33812.1; JOINED.
 DR EMBL: AB010300; BAA33812.1; JOINED.
 DR EMBL: AB010301; BAA33812.1; JOINED.
 DR EMBL: AB010302; BAA33812.1; JOINED.
 DR EMBL: AB010303; BAA33812.1; JOINED.
 DR EMBL: AB010304; BAA33812.1; JOINED.
 DR EMBL: AB010305; BAA33812.1; JOINED.
 DR EMBL: AB010306; BAA33812.1; JOINED.
 DR EMBL: AB010307; BAA33812.1; JOINED.
 DR EMBL: AB010308; BAA33812.1; JOINED.
 DR EMBL: AB010309; BAA33812.1; JOINED.
 DR EMBL: AB010310; BAA33812.1; JOINED.
 DR EMBL: AB010311; BAA33812.1; JOINED.
 DR EMBL: AB010312; BAA33812.1; JOINED.
 DR EMBL: AB010313; BAA33812.1; JOINED.
 DR EMBL: AB010314; BAA33812.1; JOINED.
 DR EMBL: AB010315; BAA33812.1; JOINED.
 DR EMBL: AB010316; BAA33812.1; JOINED.
 DR EMBL: AB010317; BAA33812.1; JOINED.
 DR EMBL: AB010318; BAA33812.1; JOINED.
 DR EMBL: AB010319; BAA33812.1; JOINED.
 DR EMBL: AB010320; BAA33812.1; JOINED.
 DR EMBL: AB010321; BAA33812.1; JOINED.
 DR EMBL: AB010322; BAA33812.1; JOINED.
 DR EMBL: AB010323; BAA33812.1; JOINED.
 DR EMBL: AB010324; BAA33812.1; JOINED.
 DR EMBL: AB010325; BAA33812.1; JOINED.
 DR EMBL: AB010326; BAA33812.1; JOINED.
 DR EMBL: AB010327; BAA33812.1; JOINED.
 DR EMBL: AB010328; BAA33812.1; JOINED.
 DR EMBL: AB010329; BAA33812.1; JOINED.
 DR EMBL: AB010330; BAA33812.1; JOINED.
 DR EMBL: AB010331; BAA33812.1; JOINED.
 DR EMBL: AB010332; BAA33812.1; JOINED.
 DR EMBL: AB010333; BAA33812.1; JOINED.
 DR EMBL: AB010334; BAA33812.1; JOINED.
 DR EMBL: AB010335; BAA33812.1; JOINED.
 DR EMBL: AB010336; BAA33812.1; JOINED.
 DR EMBL: AB010337; BAA33812.1; JOINED.
 DR EMBL: AB010338; BAA33812.1; JOINED.
 DR EMBL: AB010339; BAA33812.1; JOINED.
 DR EMBL: AB010340; BAA33812.1; JOINED.
 DR EMBL: AB010341; BAA33812.1; JOINED.
 DR EMBL: AB010342; BAA33812.1; JOINED.
 DR EMBL: AB010343; BAA33812.1; JOINED.
 DR EMBL: AB010344; BAA33812.1; JOINED.
 DR EMBL: AB010345; BAA33812.1; JOINED.
 DR EMBL: AB010346; BAA33812.1; JOINED.
 DR EMBL: AB010347; BAA33812.1; JOINED.
 DR EMBL: AB010348; BAA33812.1; JOINED.
 DR EMBL: AB010349; BAA33812.1; JOINED.
 DR EMBL: AB010350; BAA33812.1; JOINED.
 DR EMBL: AB010351; BAA33812.1; JOINED.
 DR EMBL: AB010352; BAA33812.1; JOINED.
 DR EMBL: AB010353; BAA33812.1; JOINED.
 DR EMBL: AB010354; BAA33812.1; JOINED.
 DR EMBL: AB010355; BAA33812.1; JOINED.
 DR EMBL: AB010356; BAA33812.1; JOINED.
 DR EMBL: AB010357; BAA33812.1; JOINED.
 DR EMBL: AB010358; BAA33812.1; JOINED.
 DR EMBL: AB010359; BAA33812.1; JOINED.
 DR EMBL: AB010360; BAA33812.1; JOINED.
 DR EMBL: AB010361; BAA33812.1; JOINED.
 DR EMBL: AB010362; BAA33812.1; JOINED.
 DR EMBL: AB010363; BAA33812.1; JOINED.
 DR EMBL: AB010364; BAA33812.1; JOINED.
 DR EMBL: AB010365; BAA33812.1; JOINED.
 DR EMBL: AB010366; BAA33812.1; JOINED.
 DR EMBL: AB010367; BAA33812.1; JOINED.
 DR EMBL: AB010368; BAA33812.1; JOINED.
 DR EMBL: AB010369; BAA33812.1; JOINED.
 DR EMBL: AB010370; BAA33812.1; JOINED.
 DR EMBL: AB010371; BAA33812.1; JOINED.
 DR EMBL: AB010372; BAA33812.1; JOINED.
 DR EMBL: AB010373; BAA33812.1; JOINED.
 DR EMBL: AB010374; BAA33812.1; JOINED.
 DR EMBL: AB010375; BAA33812.1; JOINED.
 DR EMBL: AB010376; BAA33812.1; JOINED.
 DR EMBL: AB010377; BAA33812.1; JOINED.
 DR EMBL: AB010378; BAA33812.1; JOINED.
 DR EMBL: AB010379; BAA33812.1; JOINED.
 DR EMBL: AB010380; BAA33812.1; JOINED.
 DR EMBL: AB010381; BAA33812.1; JOINED.
 DR EMBL: AB010382; BAA33812.1; JOINED.
 DR EMBL: AB010383; BAA33812.1; JOINED.
 DR EMBL: AB010384; BAA33812.1; JOINED.
 DR EMBL: AB010385; BAA33812.1; JOINED.
 DR EMBL: AB010386; BAA33812.1; JOINED.
 DR EMBL: AB010387; BAA33812.1; JOINED.
 DR EMBL: AB010388; BAA33812.1; JOINED.
 DR EMBL: AB010389; BAA33812.1; JOINED.
 DR EMBL: AB010390; BAA33812.1; JOINED.
 DR EMBL: AB010391; BAA33812.1; JOINED.
 DR EMBL: AB010392; BAA33812.1; JOINED.
 DR EMBL: AB010393; BAA33812.1; JOINED.
 DR EMBL: AB010394; BAA33812.1; JOINED.
 DR EMBL: AB010395; BAA33812.1; JOINED.
 DR EMBL: AB010396; BAA33812.1; JOINED.
 DR EMBL: AB010397; BAA33812.1; JOINED.
 DR EMBL: AB010398; BAA33812.1; JOINED.
 DR EMBL: AB010399; BAA33812.1; JOINED.
 DR EMBL: AB010400; BAA33812.1; JOINED.
 DR EMBL: AB010401; BAA33812.1; JOINED.
 DR EMBL: AB010402; BAA33812.1; JOINED.
 DR EMBL: AB010403; BAA33812.1; JOINED.
 DR EMBL: AB010404; BAA33812.1; JOINED.
 DR EMBL: AB010405; BAA33812.1; JOINED.
 DR EMBL: AB010406; BAA33812.1; JOINED.
 DR EMBL: AB010407; BAA33812.1; JOINED.
 DR EMBL: AB010408; BAA33812.1; JOINED.
 DR EMBL: AB010409; BAA33812.1; JOINED.
 DR EMBL: AB010410; BAA33812.1; JOINED.
 DR EMBL: AB010411; BAA3381

Db 61 SGYSQSYGGEYENOKOSYSQPPYNNQGGQQNNMESSGSGGGRAPSYDQDPY-GQQDSYDQ 119
Qy 108 QGSGGAFSSSLASMAOSYLGCGOTOQSNQOQYNNQGGNNQOYQOQGGANYQHQQGGQQQQQ 167
Db 120 SGY-----DQHGSYDEQSNYDQHDYSQNOQSYHSQRENYSHHTQDDRD- 166
Qy 168 GHSSSFSLASMASSYLGNNNSNSSSYGGQQQANEX--GRPOHNGQQQSNEXGRPOYGG 224
Db 167 -----VSRYGEDNRRGYGSGGGRGRGYDKDGRGPMTGSSGGDRGCFKNFGG 214
Qy 225 NQN-----SNGHESFNFSGN 240
Db 215 HRDYGPRTDADSESDNSDN 234

Search completed: February 1, 2002, 15:07:33
Job time: 313 sec

Db 333 SSYLGNSSNSSSYGGQQQANEXGRPOHNGQQSSNEXGRPOYGGNONGCHESFNFSGN 392
 QY 241 FSOONNNGNONRY 253
 Db 393 FSOONNNGNONRY 405

RESULT 2
 T22330
 hypothetical protein F47A4.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
 C:Accession: T22330
 R:Moritome, B.

submitted to the EMBL Data Library, June 1995

A:Reference number: Z19549

A:Accession: T22330

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-3498 <MIL>

A:Cross-references: EMBL:Z49888; PIDN:CAA90064.1; GSPDB:GN00028; CESP:F47A4.2

A:Experimental source: clone F47A4

C:Genetics:

A:Gene: CESP:F47A4.2

A:Map position: X

A:Introns: 36/3; 72/3; 147/2; 228/3; 261/2; 320/2; 847/3; 881/1; 1317/2; 1500/3; 1653/2;

Query Match 17.7%; Score 239; DB 2; Length 3498;
 Best Local Similarity 31.4%; Pred. No. 1.9e-08;

Matches 101; Conservative 20; Mismatches 117; Indels 84; Gaps 17;

QY 1 QGQGG-QGGGGGQGGGGGQGGGSGFPLALSLASSFMNSNNNNQGG----- 43
 Db 3033 QGGSTGQNVPGTQGGGGGGGGSTPQ-AQQQQPNQYSSNQYGGQAQQQQQPLNN 3091
 QY 44 -NOSGGSGFALASSFMNSNNNNNNNSOQYNSQYONGNSQYNNQYOGGNG 102
 Db 3092 VSGQSQAAGGR-PSQDSAVQSGYNTG-----NQSYPQDQDQGAQQNMSSSNQ 3143
 QY 103 GYQQ-----QGGSGGAFSSLAS-----MAQS--YLG-GGQTOSNQOYNOGQ 143
 Db 3144 AQNQLRSQQQAQPLQPPQSOQFOQPAQQAQKPMASQAQYGGGQGGQDQ--QQGQ 3201
 QY 144 NNOQQYQ--QQGQNYQHQGGQQGGHSSSFALASSMSS-----YLGN 186
 Db 3202 IAPQQAQNPQASQY-GQQDTQNRIGWSSGYTANSGSSNILLNQSMESSGLNQGSQA 3260
 QY 187 NSNSNSSYGG--GQAANEYGRPOHNGQQSSNEXGRPOY-----GNONSNGCHES--- 234
 Db 3261 SSNASSQGGSSSQMQSGYGMPCNQMQNQNQYQGRAPTMGQTNMKGSGMGSGMG 3320
 QY 235 -----FNFSGNFSQNNNGNQ 250
 Db 3321 QTGMSRSLGGLGGIGQGGQGGSQ 3342

RESULT 3
 T23339
 hypothetical protein F21C10.8 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T23339

R:Du, Z.; Galtung, S.

submitted to the EMBL Data Library, April 1996

A:Description: The sequence of C. elegans cosmid F21C10.

A:Reference number: Z20610

A:Accession: T23339

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-362 <DUZ>

A:Cross-references: EMBL:U55364; PIDN:AAA97972.1; GSPDB:GN00023; CESP:F21C10.8

A:Experimental source: strain Bristol N2; clone F21C10
 C:Genetics:
 A:Gene: CESP:F21C10.8
 A:Map position: 5
 A:Introns: 20/1; 219/3; 249/3

Query Match 17.6%; Score 237; DB 2; Length 382;
 Best Local Similarity 28.7%; Pred. No. 3e-09;
 Matches 83; Conservative 37; Mismatches 92; Indels 82; Gaps 13;

QY 15 QGQGGSFALASSFMNSNNNNQGGQSSGSGFALASSFMNSNNNSNS 74
 Db 73 QNOQYGS-----TAQYNQMNQMNQNSNQYSTSNW-----NNQNSNQY 112
 QY 75 QGQYNSY-----QNGNS-----QYNNQYQGGNGYQOQGGSGAFSSLSMA 122
 Db 113 QGQGNQNYGTQAOYDAQONQSGQYQYSSNNGYSNQNSQMSQSSQSSG---SSVSILM 169
 QY 123 QSYLGGGQTSNQOYNOQGNQ--QYQOQGNQYHQ--QGGQQQQGH-----SSSFS 174
 Db 170 DYSFNNNGCQKQGVYENQTRQATQDELQIQYRNSVQYNNQYNGYGMINSYFQ 229
 QY 175 ALASSASY-----LGNSSNS-----SYGQ-----QQ 199
 Db 230 TLPSVNSFPNIPMPMPAPCLCNAQNGCNTTNTAIGTPORDAYGNQMGSSNQFOQ 289
 QY 200 ANEXGRQHNQ--QOQSSNEXGRPOYGGNONSNGCHESFNFSQNNN 247
 Db 290 ANQGGYQNTQYQOQNGQNGQNP---QQSMGQNSQYQOQTSNQNNN 335

RESULT 4

T20847
 hypothetical protein F13E9.4 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T20847

R:McMurry, A.

submitted to the EMBL Data Library, February 1996

A:Reference number: Z19332

A:Accession: T20847

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-409 <MIL>

A:Cross-references: EMBL:Z69383; PIDN:CAA93412.1; GSPDB:GN00022; CESP:F13E9.4

A:Experimental source: clone F13E9

C:Genetics:

A:Gene: CESP:F13E9.4

A:Map position: 4

A:Introns: 32/1; 275/3; 337/3

Query Match 17.2%; Score 232; DB 2; Length 409;
 Best Local Similarity 27.9%; Pred. No. 6.9e-09;
 Matches 93; Conservative 29; Mismatches 121; Indels 90; Gaps 15;

QY 1 QGQGGGQGGGQGGG-----QGG-----QGGSFALASSFMNSNNNNQGG-----Q 43
 Db 66 QGYGQNGSMQGYSGQGYGNSQDDYGSQSGSGMGVQGYGGS---SQSTGQAQPAQQ 122
 QY 44 NOSGGSGFALASSFMNSNNNNNNNSOQYNSQYONGNSQYNNQYOGGNG 89
 Db 123 RPQGGFQSGNFSGQAAGSNNSFGQGGQGYGQNGQNGFGQGSF---SGSGMNSNLSANS 180
 QY 90 QGYNNQYQGGNGGY---QGGQSGGAFSSLASM-----AQSYLG 127
 Db 181 NGNNN---QSSSGYQNNQGRHQGGGSHSSSVMSNNGYSNSGYNNGPTPSFLN 237
 QY 128 GGYQNSQOQY-----NQGGNNQOQYQOQGNQYHQNOQOQ---QOQGH 170
 Db 238 NVSSAAQDYNTVNNKSLTTNQTNEQASNMASANSVQAOYQYETNRSALHTQATONVS 297

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 1, 2002, 15:02:16 ; Search time 19.74 Seconds
(without alignments)
288.416 Million cell updates/sec

Title: US-09-591-632-50_COPY_153_405

Perfect score: 1347
Sequence: 1 OGCGGCGGCGGCGGCGGCGG.....SFNFSGNFSQNNNGNQNRY 253

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/Backfill1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	198	14.7	678	5 PCT-US93-03027-3	Sequence 3, Appl
2	183.5	13.6	2703	1 US-08-185-432-19	Sequence 19, Appl
3	163.5	12.1	595	1 US-08-425-069-4	Sequence 4, Appl
4	163.5	12.1	595	2 US-08-317-844B-4	Sequence 4, Appl
5	158.5	11.8	903	2 US-08-853-310-2	Sequence 2, Appl
6	156.5	11.6	401	4 US-09-219-849-34	Sequence 34, Appl
7	155.5	11.5	529	4 US-09-247-806-2	Sequence 2, Appl
8	150	11.1	485	2 US-08-749-391-2	Sequence 2, Appl
9	150	11.1	485	4 US-09-390-200-2	Sequence 2, Appl
10	145	10.8	542	1 US-07-814-964-13	Sequence 13, Appl
11	145	10.8	542	1 US-08-258-442-13	Sequence 13, Appl
12	145	10.8	542	1 US-08-328-809-8	Sequence 13, Appl
13	145	10.8	542	5 PCT-US92-11107-13	Sequence 13, Appl
14	145	10.6	729	4 US-09-625-188-20	Sequence 20, Appl
15	142.5	10.6	493	4 US-08-556-978B-59	Sequence 59, Appl
16	141.5	10.5	714	4 US-08-556-978B-61	Sequence 61, Appl
17	141.5	10.5	714	4 US-09-247-806-10	Sequence 10, Appl
18	138.5	10.3	760	1 US-08-195-152-2	Sequence 2, Appl
19	131	9.7	668	1 US-08-468-036-5	Sequence 5, Appl
20	131	9.7	668	2 US-08-376-843-5	Sequence 5, Appl
21	130.5	9.7	431	4 US-09-416-050A-8	Sequence 8, Appl
22	130.5	9.7	431	4 US-09-664-800-8	Sequence 8, Appl
23	130.5	9.7	431	4 US-09-665-309-8	Sequence 8, Appl
24	130.5	9.7	431	4 US-09-661-569-8	Sequence 8, Appl
25	129	9.6	651	4 US-08-556-978B-19	Sequence 19, Appl
26	129	9.6	651	4 US-09-247-806-1	Sequence 1, Appl
27	129	9.6	718	1 US-08-425-069-2	Sequence 2, Appl

28	129	9.6	718	2 US-08-317-844B-2	Sequence 2, Appl
29	129	9.6	747	3 US-09-034-177-3	Sequence 3, Appl
30	126.5	9.4	341	2 US-08-538-711A-8	Sequence 8, Appl
31	126.5	9.4	341	4 US-08-725-027-8	Sequence 8, Appl
32	126.5	9.4	353	2 US-08-538-711A-7	Sequence 7, Appl
33	126.5	9.4	353	4 US-08-725-027-7	Sequence 7, Appl
34	126.5	9.4	414	1 US-08-343-682-1	Sequence 1, Appl
35	126.5	9.4	414	2 US-08-705-660-26	Sequence 26, Appl
36	126.5	9.4	414	3 US-08-989-045-26	Sequence 26, Appl
37	126	9.4	591	3 US-08-965-903B-2	Sequence 2, Appl
38	125.5	9.3	428	1 US-08-190-802A-29	Sequence 29, Appl
39	125.5	9.3	428	4 US-08-477-346-29	Sequence 29, Appl
40	125	9.3	656	2 US-08-343-443B-2	Sequence 2, Appl
41	125	9.3	656	4 US-09-214-564A-4	Sequence 4, Appl
42	125	9.3	1250	4 US-08-441-139-9	Sequence 9, Appl
43	123	9.1	606	4 US-08-556-978B-21	Sequence 21, Appl
44	123	9.1	606	4 US-09-247-806-4	Sequence 4, Appl
45	122.5	9.1	353	2 US-08-687-702-37	Sequence 37, Appl

ALIGNMENTS

RESULT 1
PCT-US93-03027-3
Sequence 3, Application PC/TUS9303027
GENERAL INFORMATION:
APPLICANT: LEONARD, WARREN; TOLEDANO,
MICHEL
TITLE OF INVENTION: CONTROL AND/OR
PREVENTION OF BINDING OF NF-B/REL/DORSAL
TITLE OF INVENTION:
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/03027
FILING DATE: 19930401
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/862,987
FILING DATE: 06-Apr-1992
ATTORNEY/AGENT INFORMATION:
NAME: DOROTHY R. AUTH
REGISTRATION NUMBER: P-36,434
REFERENCE/DOCKET NUMBER: 2026-4010 PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-758-4800
TELEFAX: 212-751-6849
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 678
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: No
ORIGINAL SOURCE:
ORGANISM: Drosophila melanogaster
STRAIN: Oregon R
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE: embryo
HAPLOTYPE:
TISSUE TYPE:

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/425,069
FILING DATE: 19-APR-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1447-106P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 595 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-425-069-4

Query Match 12.1%; Score 163.5; DB 1; Length 595;
Best Local Similarity 28.4%; Pred. No. 6e-07;
Matches 78; Conservative 11; Mismatches 101; Indels 85; Gaps 10;

OY 2 GGGGQGGG---GGGGGQGGGQGGGFTALASLASSFMNNSNNNOG-----42
DB 3 GYPPGGQGGGPGYPPGGGPGSAAAAAAGGPGYPPGGGPGGPGGPGRYG 62
OY 43 ---QNGSSGSSFGALASMASSFMHNSNNNSNOGYNGYNGNQNSQYNNQYOG 99
DB 63 PGQGGPGSGPSAAAAAG-----SGQGGPGYGGPRQ-- 94
OY 100 GNGGY-QGGGQGGGAFSSLSMAQSYLGCGTQSNQOQYNNQOQYQGGQYQH 158
DB 95 GPGGYGGGQGGPGSGPSAAAAAASAESG-----QGGPGYGGGPGGPGYGP 143
OY 159 QGGGQ---QGGGHSFSSALASMASSYLGNNSNSSYG-QGGQANEXGRPHN---- 209
DB 144 GQGGPGGYPGGGPGSGPSAAAAAAG--SGPGQGGPGGPGGPGGPGGPGSGP 202
OY 210 -----GQGGSNEXGRPYGGNNSNGQ 231
DB 203 GSAAAAAASGPGQGGPGGYPGGGPGGPGGPGG 237

RESULT 4
US-08-317-844B-4
Sequence 4, Application US/08317844B
Patent No. 5989894
GENERAL INFORMATION:
APPLICANT: Lewis, Randolph V.
APPLICANT: Xu, Ming
TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK
TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL
TITLE OF INVENTION: CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF
NUMBER OF SEQUENCES: 62
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch
STREET: 301 No. 5989894th Washington Street
CITY: Falls Church
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22046
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/317,844B
FILING DATE: 04-OCT-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1447-105P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 241-1300
TELEFAX: (703) 241-2848
TELEX: 248345
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 595 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-317-844B-4

Query Match 12.1%; Score 163.5; DB 2; Length 595;
Best Local Similarity 28.4%; Pred. No. 6e-07;
Matches 78; Conservative 11; Mismatches 101; Indels 85; Gaps 10;

OY 2 GGGGQGGG---GGGGGQGGGQGGGFTALASLASSFMNNSNNNOG-----42
DB 3 GYPPGGQGGGPGYPPGGGPGSAAAAAAGGPGYPPGGGPGGPGGPGRYG 62
OY 43 ---QNGSSGSSFGALASMASSFMHNSNNNSNOGYNGYNGNQNSQYNNQYOG 99
DB 63 PGQGGPGSGPSAAAAAG-----SGQGGPGYGGPRQ-- 94
OY 100 GNGGY-QGGGQGGGAFSSLSMAQSYLGCGTQSNQOQYNNQOQYQGGQYQH 158
DB 95 GPGGYGGGQGGPGSGPSAAAAAASAESG-----QGGPGYGGGPGGPGYGP 143
OY 159 QGGGQ---QGGGHSFSSALASMASSYLGNNSNSSYG-QGGQANEXGRPHN---- 209
DB 144 GQGGPGGYPGGGPGSGPSAAAAAAG--SGPGQGGPGGPGGPGGPGGPGSGP 202
OY 210 -----GQGGSNEXGRPYGGNNSNGQ 231
DB 203 GSAAAAAASGPGQGGPGGYPGGGPGGPGGPGG 237

RESULT 5
US-08-853-310-2
Sequence 2, Application US/0853310
Patent No. 5948640
GENERAL INFORMATION:
APPLICANT: Randazzo, Filippo
TITLE OF INVENTION: Mammalian Additional Sex Combs (Asx) Acts as a Tumor Suppre
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: California
COUNTRY: U.S.A.
ZIP: 94608
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/853,310
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Guth, Joseph H.
REGISTRATION NUMBER: 31,261
REFERENCE/DOCKET NUMBER: 1228.003

Sequence 2, Application US/08749391
Patent No. 5948667
GENERAL INFORMATION:
APPLICANT: Cheng, Kuo-Joan
APPLICANT: Selinger, Leonard B.
APPLICANT: Liu, Jin-Hao
APPLICANT: Hu, Youji
APPLICANT: Forsberg, Cecil W.
APPLICANT: Moloney, Maurice M.
TITLE OF INVENTION: A xylanase Obtained From an
TITLE OF INVENTION: Anaerobic Fungus
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
STREET: 5370 Manhattan Circle, Suite 201
CITY: Boulder
STATE: Colorado
COUNTRY: USA
ZIP: 80803
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/749,391
FILING DATE: 13-NOV-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Donna M. Feiber
REGISTRATION NUMBER: 33,878
REFERENCE/DOCKET NUMBER: 93-96
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 499-8080
TELEFAX: (303) 499-8089
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 485 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-749-391-2

Query Match 11.1%; Score 150; DB 2; Length 485;
Best Local Similarity 28.4%; Pred. No. 7.4e-06;
Matches 69; Conservative 11; Mismatches 117; Indels 46; Gaps 8;
QY 14 GGGGGSFTALASLASSFMNSNNNNQGGSSGSGFALASMASSFMHNNNNNSN 73
DB 223 GNGNGVSGTADFPYAKVYIGDNGGASPARAGAGAPAGAP-----GND 268
QY 74 SGGYNSYNGNNGNSGYNNOYQGGNGGYQQQQGSGGAFSSLASMAOSYLAGGQTQS 133
DB 269 QPQG-----PQGQGPQGQGPQGQGPQGQGPQGQGPQGQGPQGQGPQGQGP 305
QY 134 NQQQYNO--QGQNNQ--QYQQQGNVYHQQQGQQQQGSHSSFSALASMASSYLGNNNS 190
DB 306 NDDQGGQPPQGQGPQGQNDQDQGGQ--PPQGPQGNGPGGSDFFNNWQGGSPWGGN--QG 362
QY 191 NSSYGGQQQANEYGRPHNGQQQGSNEYGRPOYGGNNSN--GQHSFNFSGNFSQQNNN 247
DB 363 GSPWGGNGGNGPQNGGSGSPWGGNGGSPWGGNGGNGGNGGNGGNGGNGGNGGNGP 422
QY 248 GNG 250
DB 423 GNG 425

RESULT 9
US-09-390-200-2
Sequence 2, Application US/09390200
Patent No. 6137032

GENERAL INFORMATION:
APPLICANT: Cheng, Kuo-Joan
APPLICANT: Selinger, Leonard B.
APPLICANT: Liu, Jin-Hao
APPLICANT: Hu, Youji
APPLICANT: Forsberg, Cecil W.
APPLICANT: Moloney, Maurice M.
TITLE OF INVENTION: A xylanase Obtained From an
TITLE OF INVENTION: Anaerobic Fungus
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
STREET: 5370 Manhattan Circle, Suite 201
CITY: Boulder
STATE: Colorado
COUNTRY: USA
ZIP: 80803
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/390,200
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/749,391
FILING DATE: 13-NOV-1996
ATTORNEY/AGENT INFORMATION:
NAME: Donna M. Feiber
REGISTRATION NUMBER: 33,878
REFERENCE/DOCKET NUMBER: 93-96
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 499-8080
TELEFAX: (303) 499-8089
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 485 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-390-200-2

Query Match 11.1%; Score 150; DB 4; Length 485;
Best Local Similarity 28.4%; Pred. No. 7.4e-06;
Matches 69; Conservative 11; Mismatches 117; Indels 46; Gaps 8;
QY 14 GGGGGSFTALASLASSFMNSNNNNQGGSSGSGFALASMASSFMHNNNNNSN 73
DB 223 GNGNGVSGTADFPYAKVYIGDNGGASPARAGAGAPAGAP-----GND 268
QY 74 SGGYNSYNGNNGNSGYNNOYQGGNGGYQQQQGSGGAFSSLASMAOSYLAGGQTQS 133
DB 269 QPQG-----PQGQGPQGQGPQGQGPQGQGPQGQGPQGQGPQGQGPQGQGP 305
QY 134 NQQQYNO--QGQNNQ--QYQQQGNVYHQQQGQQQQGSHSSFSALASMASSYLGNNNS 190
DB 306 NDDQGGQPPQGQGPQGQNDQDQGGQ--PPQGPQGNGPGGSDFFNNWQGGSPWGGN--QG 362
QY 191 NSSYGGQQQANEYGRPHNGQQQGSNEYGRPOYGGNNSN--GQHSFNFSGNFSQQNNN 247
DB 363 GSPWGGNGGNGPQNGGSGSPWGGNGGSPWGGNGGNGGNGGNGGNGGNGGNGGNGP 422
QY 248 GNG 250
DB 423 GNG 425

RESULT 10
US-07-814-964-13
Sequence 13, Application US/07814964


```
Cy      37 NNNDGGGNOSSGGSFFGALASMAS-----SEF-----63
          :|::|||:|||||:
Db       69 SSSNNNNVYVALPDQFGMOQTGGSNYPYINQQFSEFYNSFLSHLTKOTNPSTVTGTG 128
Cy      64 -HSNNNONSNNSGGYNGSYONGNQ-----NSOGYNNDQYGNGGYOQQOQG- 110
          ||||:|:||:|||||:|||||:
Db     129 ASSNNNSSNNNVSSGNNSTSSNPQLAASOLNPATATTAAANN---AAGPASTYSLPQY 165
Cy     111 -----SGAFSSL--ASMAOSTYIGGCOTOS-----NQOYTNDGGGNMQ 147
          ::|||:|||||:|||||:
Db     186 QRYRPPNNNAALSLIDPSSACNAGNANTATHPLCLPPNLDPOLTHHQDDMOODLOLDQ 245
Cy     148 QYQGGQNGNYQHQQGGGOQQO 167
          |::|||:|||||:
Db    246 QQDLDQQQQDLQQQHDLDQQQQ 265
```

```

RESULT 14
US-09-625-188-20
: Sequence 20, Application US/09625188
: Patent No. 6307037
: GENERAL INFORMATION:
: APPLICANT: No. 6307037artis AG
: TITLE OF INVENTION: Fungal Target Genes and Methods
: FILE REFERENCE: PB/5-1126P1
: CURRENT APPLICATION NUMBER: US/09/625.188
: CURRENT FILING DATE: 2000-07-21
: NUMBER OF SEQ ID NOS: 44
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 20
: LENGTH: 729
: TYPE: PRF
: ORGANISM: Asibhya gossypii
: US-09-625-188-20

```

	Query Match	10.8%	Score 145;	DB 4;	Length 729;	
	Best Local Similarity	25.2%;	Pred. No. 3.4e-05;			
	Matches	62;	Conservative	26;	Mismatches 114;	Indels 44; Gaps 6
QY	16	GCGGCFPALALSLASSFPMNSNN--NQCGONDSGGSGFGALASMASSFMSHNNNSNN	73			
Db	340	NGGTDPISST::: :				
QY	74	SQCGYNQSYQNQNSNQGYN-----NQCYGNGGYQQQOGSGAGFSLASMAQS	124			
Db	398	LATGIQQLPLPQRQRLSLNLVTPLSLSDMNLPRSSQGGILRHQAQCQPQSQALQHHQ	457			
QY	125	YLGGGQTQSNQOQYNQCGGNNDQOYQQGQANTQHQQGQGGQQGSHSSSFSLASMASTYL	184			
Db	458	HLHHQOQDLQOQOHLHQOQHQQ--QOOSLSQOPQOQSOQSOQASHSQO-----	503			
QY	185	GNNSNSNSSYGQOQDAEXGRPOHNGQOOSENYGRPYQYGGNNSNQHESFNFSNFSSQ	244			
Db	504	-----HGQGHQOQOQOQPPQDQRPQRRHPQRPQOQONSQDA-----IYGSQQQ	546			
QY	245	NNNNGQ 250				
Db	547	VTSQQQ 552				

RESULT 15
US-08-556-978B-59
; Sequence 59, Application US/08556978B
; Patent No. 6268169
; GENERAL INFORMATION:
; APPLICANT: FAHNSTOCK, STEPHEN F.
; TITLE OF INVENTION: NOVEL RECOMBINANTLY PRODUCED
; TITLE OF INVENTION: SPIDER SILK ANALOGS

```

1 NUMBER OF SEQUENCES: 107
2 CORRESPONDENCE ADDRESS:
3 ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
4 STREET: 1007 MARKET STREET
5 CITY: WILMINTON
6 STATE: DELAWARE
7 COUNTRY: UNITED STATES OF AMERICA
8 ZIP: 19898
9
10 COMPUTER READABLE FORM:
11 MEDIUM TYPE: DISKETTE, 3.50 INCH
12 COMPUTER: IBM PC COMPATIBLE
13 OPERATING SYSTEM: MICROSOFT WINDOWS 95
14 SOFTWARE: MICROSOFT WORD FOR WINDOWS 95
15
16 CURRENT APPLICATION DATA:
17 APPLICATION NUMBER: US/08/556,978B
18 FILING DATE:
19 CLASSIFICATION: 435
20
21 PRIOR APPLICATION DATA:
22 APPLICATION NUMBER: 08/077,600
23 FILING DATE: JUNE 15, 1993
24
25 ATTORNEY/AGENT INFORMATION:
26 NAME: FLOYD, LINDA AXAMETHY
27 REGISTRATION NUMBER: 33,692
28 REFERENCE/DOCKET NUMBER: CR-9389-A
29
30 TELECOMMUNICATION INFORMATION:
31 TELEPHONE: 302-892-8112
32 TELEFAX: 302-773-0164
33
34 INFORMATION FOR SEQ ID NO: 59:
35 SEQUENCE CHARACTERISTICS:
36 LENGTH: 493 amino acids
37 TYPE: amino acid
38 STRANDEDNESS: unknown
39 TOPOLOGY: unknown
40
41 MOLECULE TYPE: protein
42
43 US-08-556-978B-59

```

[illegible]

Search completed: February 1, 2002, 15:02:17
Job time: 237 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compen Ltd.

OM protein - protein search, using sw model

Run on: February 1, 2002, 15:01:49 ; Search time 38.77 Seconds
(without alignments)
483.377 Million cell updates/sec

Title: US-09-591-632-50_COPY_153_405

Perfect score: 1347
Sequence: 1 QGQGQGGGQGGGQGGGQGGG.....SFNFGNFSQNNNGNNGNRY 253

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq_1101.*
1: /SID8/gcgdata/geneseq/geneseqp/AA1980.DAT.*
2: /SID8/gcgdata/geneseq/geneseqp/AA1981.DAT.*
3: /SID8/gcgdata/geneseq/geneseqp/AA1982.DAT.*
4: /SID8/gcgdata/geneseq/geneseqp/AA1983.DAT.*
5: /SID8/gcgdata/geneseq/geneseqp/AA1984.DAT.*
6: /SID8/gcgdata/geneseq/geneseqp/AA1985.DAT.*
7: /SID8/gcgdata/geneseq/geneseqp/AA1986.DAT.*
8: /SID8/gcgdata/geneseq/geneseqp/AA1987.DAT.*
9: /SID8/gcgdata/geneseq/geneseqp/AA1988.DAT.*
10: /SID8/gcgdata/geneseq/geneseqp/AA1989.DAT.*
11: /SID8/gcgdata/geneseq/geneseqp/AA1990.DAT.*
12: /SID8/gcgdata/geneseq/geneseqp/AA1991.DAT.*
13: /SID8/gcgdata/geneseq/geneseqp/AA1992.DAT.*
14: /SID8/gcgdata/geneseq/geneseqp/AA1993.DAT.*
15: /SID8/gcgdata/geneseq/geneseqp/AA1994.DAT.*
16: /SID8/gcgdata/geneseq/geneseqp/AA1995.DAT.*
17: /SID8/gcgdata/geneseq/geneseqp/AA1996.DAT.*
18: /SID8/gcgdata/geneseq/geneseqp/AA1997.DAT.*
19: /SID8/gcgdata/geneseq/geneseqp/AA1998.DAT.*
20: /SID8/gcgdata/geneseq/geneseqp/AA1999.DAT.*
21: /SID8/gcgdata/geneseq/geneseqp/AA2000.DAT.*
22: /SID8/gcgdata/geneseq/geneseqp/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1347	100.0	405	22	AAB30810
2	1347	100.0	405	22	AAB30821
3	895	66.4	407	10	AA94879
4	203.5	15.1	461	16	AA848550
5	198.5	14.7	280	21	AA863197
6	198	14.7	678	14	AA842087
7	192	14.3	342	17	AAW03625
8	189.5	14.1	145	22	AA869614
9	187.5	13.9	894	22	AA835148
10	187.5	13.9	271	22	AA830800
11	184.5	13.7			A modified N regio

12	184.5	13.7	437	16	AA875507
13	184.5	13.7	729	22	AA670847
14	183.5	13.6	590	12	AA811019
15	179.5	13.3	155	22	AA869612
16	178	13.2	113	22	AA869615
17	175	13.0	256	13	AA824796
18	175	13.0	724	21	AA831571
19	175	13.0	826	21	AA831570
20	173.5	12.9	404	17	AA890675
21	173.5	12.9	326	21	AA837007
22	171.5	12.7	342	21	AA837006
23	171.5	12.7	343	21	AA837005
24	171.5	12.7	345	21	AA837005
25	169.5	12.6	149	22	AA872673
26	169	12.5	537	7	AA860452
27	168.5	12.5	685	22	AA830792
28	165	12.2	477	22	AA830815
29	164	12.2	715	22	AA830820
30	163.5	12.1	531	20	AA840098
31	163.5	12.1	595	12	AA814309
32	163.5	12.1	595	12	AA853347
33	163.5	12.1	595	19	AA859071
34	162.5	12.1	767	11	AA808401
35	162.5	12.1	767	11	AA807599
36	162.5	12.1	2074	21	AA854319
37	161.5	12.0	121	22	AA869609
38	158.5	11.8	903	18	AA837389
39	158.5	11.8	903	18	AA837391
40	156.5	11.6	401	21	AA806023
41	156.5	11.6	599	22	AA872374
42	156.5	11.6	599	22	AA872375
43	155.5	11.5	3190	22	AA884634
44	155.5	11.5	409	22	AA870829
45	154.5	11.5			

ALIGNMENTS

RESULT 1
ID AAB30810 standard: Protein: 405 AA.
XX AAB30810:
XX
XX 02-APR-2001 (first entry)
XX
XX Amino acid sequence of a prion-like amyloidogenic protein.
XX
XX SCHAG: self-coalesce; higher-order aggregate; amyloidogenic domain;
XX aggregation: fibril; phenotypic alteration; gene therapy;
XX disease resistance; plant pigmentation; prion disease.
XX
XX Saccharomyces cerevisiae.
XX
XX WO200075324-A2.
XX
XX 14-DEC-2000.
XX
XX 09-JUN-2000: 2000WO-US15876.
XX
XX 09-JUN-1999: 99US-0138833.
XX
XX (ARCH-) ARCH DEV CORP.
XX
XX Lindquist S, Li L, Ma J, Liu J, Sondheimer N, Scheibel T;
XX
XX WPI: 2001-061723/07.
XX
XX New nucleic acid encoding chimeric proteins with self-assembly
XX properties, useful e.g. for diagnosis and treatment of prion diseases,
XX also related aggregates, fibrils and polymers -
XX

Nicotiana plumbagi
C albicans apoptos
Apo-11proteins E
Huntingtin accuml
Sequence of fibron
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Human chromosome-1
Arabidopsis thalia
Arabidopsis thalia
Polyglutamine trac
Sequence of the As
Amino acid sequenc
Amino acid sequenc
Amino acid sequenc
Spider silk protel
N. clavipes draglin
Nephila clavipes s
N. clavipes spider
Amf105 encoded by
Sequence deduced by
Amino acid sequenc
Huntingtin accuml
Human additional s
Human additional s
Polar gelatin P te
Amphiphilic recomb
Amphiphilic recomb
Plasmodium falcipa
Amino acid sequenc
C albicans apoptos

Db 700 -----sgggggggsgsgsygygghynspvpkhaqkxphgqgkpsysgy 749

Qy 105 000GSGSGAFSLASMAQS-YLGGGQTQSNQOQYN--OQGNQOQYQOQ-----GQNYQ 157

Db 750 gshgqgq-----qsyngspysnygppgkqkqynhggysysnsynspgpgsgsdyn 802

Qy 158 HQOQGOOQOQHSSSFSALASMASSYLGNNSNSNSSYGGOQOANEYGRPOHNGOQOSNEY 217

Db 803 yeaefnygsgrsgnsgnysgsgaay---npqshggygsgsgsgsgsgsgsgsn-y 858

Qy 218 GRPOYGGNONSNGOHESENFES-GNFSQOQNNNGNQRX 253

Db 859 nsp--gsqgnygsppsgysgsqgygrnadhemygy 893

RESULT 11

AAB30800

ID AAB30800 standard; Protein; 271 AA.

XX AAB30800;

AC

XX

DT 02-APR-2001 (first entry)

DE A modified N region of yeast Sup35 protein.

XX

XX Sup35; phenotypic; SCHAG; self-coalesce; higher-order aggregate;

XX amyloidogenic domain; aggregation; fibril; phenotypic alteration;

KW gene therapy; disease resistance; plant pigmentation; prion disease.

XX

OS Synthetic.

OS Saccharomyces cerevisiae.

XX

PN WO200075324-A2.

XX

PD 14-DEC-2000.

XX

PF 09-JUN-2000; 2000WO-US15876.

XX

PR 09-JUN-1999; 99US-0138833.

PA (ARCH-) ARCH DEV CORP.

PI Lindquist S, Li L, Ma J, Liu J, Sondheimer N, Scheibel T;

XX

DR MPI: 2001-061723/07.

XX

DR N-PSDB; AAC86685.

XX

PT New nucleic acid encoding chimeric proteins with self-assembly

PT properties, useful e.g. for diagnosis and treatment of prion diseases,

PT also related aggregates, fibrils and polymers -

XX

PS Claim 11; Page 135-136; 188pp; English.

XX

XX The present sequence represents a modified N region of Sup35 protein, in

CC the second oligopeptide repeat has been expanded twice, creating a total

CC of seven repeats. Sup35 possesses the prion-like capacity to undergo a

CC self-perpetuating conformational alteration that changes the functional

CC state of Sup35 in a manner that creates a heritable change in phenotype.

CC It is used to construct chimeric polypeptides of the invention, which

CC comprise at least one SCHAG (self-coalesces into higher-order aggregates)

CC amino acid sequence fused in frame with a polypeptide of interest (which

CC is other than a marker protein, a glutathione-S-transferase or a

CC staphylococcal nuclear protein). The specification also describes

CC chimeric polypeptides that comprises an amyloidogenic domain that causes

CC aggregation into fibrils. The chimeric polypeptides are used to prepare

CC polymers with multiple reactivities, e.g. derivatised with enzymes, or

CC specific binding partners, and useful e.g. for performing multi-step

CC chemical reactions. They can be used create an inducible, or stable

CC phenotypic alteration in a cell, e.g. for gene therapy, protein

CC production, imparting disease resistance to plants, altering plant

CC pigmentation and for diagnosis and treatment of prion diseases.

XX

SQ Sequence 271 AA:

Query Match 13.7%; Score 184.5; DB 22; Length 271;

Best Local Similarity 31.3%; Pred. No. 5.8e-08;

Matches 63; Conservative 13; Mismatches 56; Indels 69; Gaps 10;

Qy 67 NNONSNNSOQGYNQSYONGNONS-----QGYN--NOOYOGGNGCYOQOQSGSGAFSLA 119

Db 2 sdnngnngnqnyqyqgngnqgqgmnrqyqyqaynaqapqggyqnyqy99----- 54

Qy 120 SMAQSLGSGGQTQSNQOQYNQOQGNQOQYQOQOQNTQHOOQGOQOQOQHSSSFSALASW 179

Db 55 -----Yqggg-----Yqgynpqg--Ygyqynpqgyqgynpdagyqgq----- 90

Qy 180 ASSYLGNNSNSNSSYGGOQOANEYGRPOHNGOQOSNEYGRPOYGGNONSNGOHESENFSG 239

Db 91 -----Ynpqgyqqyn-----pqggyqqfn-----pqg-----rgnyknfnym 126

Qy 240 N-----FSOQNNNGNQRX 253

Db 127 nlqgyqagfgpqsqgmslndf 147

RESULT 12

AAR75507

ID AAR75507 standard; Protein; 437 AA.

XX AAR75507;

AC

XX

DT 05-DEC-1995 (first entry)

XX

DE Nicotiana plumbaginifolia arabinogalactan protein NPAGP1.

XX

KW Arabinogalactan protein; AGP; gum; thickening agent; emulsifier;

KW adhesive.

XX

OS Nicotiana plumbaginifolia.

XX

FH Key Location/Qualifiers

FT Peptide 14..24

FT Peptide /note="corresp. to microsequenced peptide"

FT 14..24

FT /note="Identical with microsequenced peptide"

FT Misc-difference 28

FT /note="see above"

FT Misc-difference 30

FT /note="see above"

FT Misc-difference 32..37

FT /note="see above"

FT Misc-difference 25

FT /label="hydroxyproline

FT Misc-difference 27

FT /label="see above

FT Misc-difference 29

FT /label="see above

FT Misc-difference 30

FT /label="see above

FT Misc-difference 26

FT /note="T in microsequenced peptide"

PN WO9515377-A.

XX

PD 08-JUN-1995.

XX

PF 01-DEC-1994; 94WO-AU00744.

XX

PR 18-JUL-1994; 94US-0276452.

XX

PR 03-DEC-1993; 93US-0161944.

XX

PA (ALBR) ALBRIGHT & WILSON AUSTRALIA.

XX

PI Bacic A, Chen C, Clarke AE, Du H, Ganu AM, Mau S;

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 1, 2002, 15:01:55 ; Search time 37.66 Seconds

(without alignments)
477.735 Million cell updates/sec

Title: US-09-591-632-2_COPY_1_123

Perfect score: 701

Sequence: 1 MSDSNQGNNQONTQOVSQNG.....FNNYNNLQGYQAGFPQSQG 123

Scoring table: BLOSUM62

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: SP:REMBL_17:*
2: SP:archaea:*
3: SP:bacteria:*
4: SP:fungi:*
5: SP:human:*
6: SP:invertebrate:*
7: SP:mammal:*
8: SP:mhc:*
9: SP:organelle:*
10: SP:phage:*
11: SP:plant:*
12: SP:rodent:*
13: SP:virus:*
14: SP:vertebrate:*
15: SP:unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	701	100.0	685	3	Q9HGV1
2	695	99.1	224	3	Q07089
3	229.5	32.7	712	3	Q9HGI7
4	228	32.5	299	3	Q9UVK3
5	212	30.2	428	3	Q9URS3
6	212	30.2	700	3	Q9HGI8
7	204	29.1	320	3	Q9UVK2
8	194.5	27.7	701	3	Q9HGI6
9	190	27.1	662	3	Q9HGI4
10	177	25.2	274	3	Q9UVJ9
11	175	25.0	742	3	Q9HGI5
12	170	24.3	200	11	Q9D036
13	167.5	23.9	305	3	Q9UVK0
14	162	23.1	2382	5	Q9B119
15	160	22.8	407	3	Q94079
16	152.5	21.8	215	3	Q43117
17	149	21.3	456	5	Q20936
18	148	21.1	419	10	Q49429
19	147.5	21.0	409	5	Q19414

20	147.5	21.0	924	5	Q20189	Q20189 caenorhabd1
21	146	20.8	497	5	Q27512	Q27512 caenorhabd1
22	144	20.5	222	5	Q20013	Q20013 caenorhabd1
23	143	20.4	584	5	Q9NJ80	Q9NJ80 spodoptera
24	141.5	20.2	492	4	Q9NRD7	Q9NRD7 homo sapien
25	141	20.1	800	13	Q91551	Q91551 xenopus lae
26	140.5	20.0	236	3	Q94060	Q94060 candida alb
27	139.5	19.9	422	4	Q9BXC6	Q9BXC6 homo sapien
28	139.5	19.9	492	4	Q9HBC2	Q9HBC2 homo sapien
29	139.5	19.9	716	3	Q42787	Q42787 podospora a
30	139	19.8	632	10	Q82762	Q82762 arabidopsis
31	138	19.7	1111	2	Q86522	Q86522 streptomyc
32	137.5	19.6	284	5	Q21073	Q21073 caenorhabd1
33	137.5	19.6	765	5	Q17796	Q17796 caenorhabd1
34	136.5	19.5	461	10	Q41256	Q41256 nicotiana a
35	135	19.3	528	4	Q13344	Q13344 homo sapien
36	135	19.3	867	5	Q18218	Q18218 caenorhabd1
37	135	19.3	871	5	Q62486	Q62486 caenorhabd1
38	134.5	19.2	319	10	Q41725	Q41725 zinnia eleg
39	133.5	19.0	254	5	Q9U7P0	Q9U7P0 eufoiliacul1
40	133	19.0	369	4	Q9HBC1	Q9HBC1 homo sapien
41	133	19.0	853	5	Q94464	Q94464 dictyostell
42	133	19.0	3498	5	Q20497	Q20497 caenorhabd1
43	131	18.7	1118	5	Q9TVH5	Q9TVH5 plasmodium
44	131	18.7	1161	5	Q9TV17	Q9TV17 plasmodium
45	131	18.7	1961	3	Q9UVK9	Q9UVK9 yarrowia 11

ALIGNMENTS

RESULT 1
ID Q9HGV1 PRELIMINARY; PRT; 685 AA.
AC Q9HGV1;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE SUP35 ALLOSUPPRESSOR MUTANT SAL3-4.
GN SUP35.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BSCT83/4A SAL3-4;
RA Resende C.G., Duarte J.B., Tuile M.F.;
RT "sal3-4 a allosuppressor mutation of the SUP35 gene.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO GTP-BINDING ELONGATION FACTOR FAMILY.
DR EMBL: AF263099; AAF9684.1; -;
DR InterPro: IPR000795; GTP_EFTU.
DR Pfam: PF00009; GTP_EFTU; 1.
DR PRINTS: PR00315; ELONGATNFACT.
KW GTP-binding; Protein biosynthesis.
SQ SEQUENCE 685 AA; 76550 MW; F3958EC37D750154 CRC64;

Query Match 100.0%; Score 701; DB 3; Length 685;
Best Local Similarity 100.0%; Pred. No. 9.3e-61;
Matches 123; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSDSNQGNNQONTQOVSQNGNNRQGYQAYNNAOAPAGGYONYGSGYQOQGY 60
Db 1 MSDSNQGNNQONTQOVSQNGNNRQGYQAYNNAOAPAGGYONYGSGYQOQGY 60
QY 61 QQYNPDAGYQOQYNNPQGYQOQYNNPQGYQOQYNNPQGYQOQYNNPQGYQOQY 120
Db 61 QQYNPDAGYQOQYNNPQGYQOQYNNPQGYQOQYNNPQGYQOQYNNPQGYQOQY 120
QY 121 SQG 123
Db 121 SQG 123

Db	121	Seq	123
RESULT	2		
Q07089			
ID	007089	PRELIMINARY;	PRT; 224 AA.
AC	007089;		
DT	01-NOV-1996 (TREMBLrel. 01, Created)		
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)		
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)		
DE	ONKIPOTENT SUPPRESSOR.		
GN	SUP35 OR YDR172W.		
OS	Saccharomyces cerevisiae (Baker's Yeast).		
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes		
CC	Saccharomycetales; Saccharomycetaceae; Saccharomyces.		
CX	NCBI_TaxID=4932.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=88329727; PubMed=3047009;		
RA	Kushnir V. V., Ter-Avanesyan M.D., Telckov M.V., Surguchov A.P		
RA	Smulnikov V.N., Inge-Vechtomov S.G.;		
RT	"Nucleotide sequence of the SUP2 (SUP35) gene of Saccharomyces		
RT	cerevisiae.";		
RL	Gene 66:55-54(1988).		
CC	-1- SIMILARITY: TO GTP-BINDING ELONGATION FACTOR FAMILY.		
DR	EMBL: M21129; AAA35133.1; .		
DR	SCD: S0002579; SUP35.		
DR	InterPro: IPR000795; GTP_EFTU.		
DR	Pfam: PF00009; GTP_EFTU; 1.		
DR	GTP-Binding; Protein biosynthesis.		
CO	SEQUENCE 224 AA; 25182 MW; 2521133DCACT7D65 CRC64;		

Query Match	99.1%	Score 695	DB 3	Length 224
Best Local Similarity	100.0%	Pred. NO. 1.1e-60		
Matches 122	Conservative 0	Mismatches 0	Indels 0	Gaps 0
Qy	1	MDSNCGNNQOYNQOYSONGNGQGNRRYOGYQAAVNAQAQAPAGGYQYQYSGYQOQGGY	60	
Db	1	MDSNCGNNQOYNQOYSONGNGQGNRRYOGYQAAVNAQAQAPAGGYQYQYSGYQOQGGY	60	
Qy	61	QOYNPDAGYQOQOYNPDQGYQOYNPDQGYQOQENPDQGRGNYKNFNYNNNLQGYQAGFQPO	120	
Db	61	QOYNPDAGYQOQOYNPDQGYQOYNPDQGYQOQENPDQGRGNYKNFNYNNNLQGYQAGFQPO	120	
Qy	121	SQ 122		
Db	121	SQ 122		
RESULT	3			
ID	Q9HG17	PRELIMINARY	PRT	712 AA.
AC	Q9HG17			
DT	01-MAR-2001 (TREMBLrel. 16, Created)			
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	POLYPEPTIDE RELEASE FACTOR 3.			
GN	SUP35.			
OS	Candida maltosa (Yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; mitosporic Saccharomycetales; Candida.			
XX	NCBI_taxid=5479;			
XX	[1]			
XX	SEQUENCE FROM N.A.			
XX	Nakavashiki T., Ebihara K., Nakamura Y.:			
XX	"Psi-like properties of fungus Sup35 proteins.";			
XX	Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.			
XX	-1- SIMILARITY: TO GTP-BINDING ELONGATION FACTOR FAMILY.			
XX	EMBL: AB039750; BAB12681.2; -			
XX	InterPro: IPR000795; GTP_EFTU.			
XX	Pfam: PF00009; GTP_EFTU, 1.			
XX	PRINTS: PR00315; ELONGATINFC.			

KW GTP-binding; Protein biosynthesis.
SQ SEQUENCE 712 AA; 78631 MW; 729B6A814735C469 CRC64;

	Query Match	32.7%	Score 229.5;	DB 3;	Length 712;
	Best Local Similarity	48.8%;	Pred. No. 1.1e-14;		
	Matches 60;	Conservative 7;	Mismatches 43;	Indels 13;	Gaps
OY	4 SNOGNNQONTYOYSO-NGNQGCGNNRKYGYOAYNAQAOPA-GGYQNYOGYSGYOQGGYQ 61				
Db	22 SKQPQQQQPQQQQPYFPNNQAQAFVPPFGYGQQFQPGQQQQYGGYQN--YTGYOAGSGY 78				
OY	62 Q-YNPDAGYQQQYTPQGSGYQ-QYNPQGSGYQQGFNPQGGRGNTYKRFNFNTNNNLQGYQACFOR 119				
Db	79 QNYNNRGYGOQNYNNRGYGOQNYNNRGYQOOOQOO-----YQAYNPNQOYGGYOA-YNP 132				
OY	120 QSQ 122				
Db	133 QQQ 135				

RESULT	4			
09UVK3				
09UVK3		PRELIMINARY;	PRT;	299 AA.
AC	09UVK3;			
DT	01-MAY-2000 (TREMBLrel. 13, Created)			
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)			
DT	01-MAY-2000 (TREMBLrel. 13, Last annotation update)			
DE	SUP35 HOMOLOG (FRAGMENT).			
GN	SUP35.			
OS	Candida albicans (Yeast).			
OC	Eukaryota: Fungi: Ascomycota: Saccharomycotina: Saccharomycetes:			
OC	Saccharomycetales: mitosporic Saccharomycetales; Candida.			
OX	NCBI_Taxid=5476;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Santosco A., Chien P., Weissman J.S.;			
RT	"Molecular Basis of a Yeast Protein Species Barrier ";			
RL	Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AF206287; AF144002.1; -.			
FT	NON_TER			
EQ	SEQUENCE 299 AA; 33702 MW; F15D7800DE73C49F CRC64;			

	Query Match	32.5%	Score 228:	DB 3:	Length 299;	
Bst Local Similarity	43.2%;	Pred. No.	6.2e-15;			
Matches 57;	Conservative 12;	Mismatches 43;		Indels 20;	Gaps 8;	
Oy	I MSDSNGNNOQNYOYSGNSGNCNNRYOGYAANAQA----PAGGY--QYNQGISG 54 I :::: : I D 15 MAMASLNGDSKSKQQQQQQQQQQNQ-----YYNPMAOSFVPGGGYGQFOFPQPQQ 67 Oy 55 YQG-GGYQYMP-DAGYQXQYPNPGGYGO-YNPQGSYGQQDFPNQGRGNKKNFNTNNNLQ 111 Db 68 QQQTGGYQMAYMORYGGYQMNRRGTYQGYNNRGTGMNTN---NRGAQSINONMQYT 124 Oy 112 GYOA-GFOPOSQ 122 Db 125 GYQQTNSOPQQ 136 RESULT 5 ID Q9URS3 PRELIMINARY; PRT; 428 AA. AC Q9URS3; DT 01-MAY-2000 (TREMBLEREL_13, created) DT 01-MAY-2000 (TREMBLEREL_13, last annotation update) DT 01-JUN-2001 (TREMBLEREL_17, last annotation update) DE SUP35 HOMOLOG (FRAGMENT). GN SUP35. OS Kluveromyces lactis (Yeast). OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; CC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.					

Db 1149 SQTVGGFQRPYSVKGGYQPHNQLNGGYNQPNQSSSFGA-NGGYQPPSQ 1195

RESULT 15

ID	094079	PRELIMINARY;	PRF:	407 AA.
AC	094079;			
ET	01-MAY-1999 (TREMBLrel. 10, Created)			
ET	01-MAY-1999 (TREMBLrel. 10, Last sequence update)			
ET	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	YEAST (S.CEREVISIAE) FUS1 (FUSION GENE) AND BIK1 GENE REGION.			
GN	RNO1 AND YCLO28M.			
OS	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
CC	Saccharomycetales; Saccharomycetaceae; Saccharomyces.			
CX	NCBI_TaxID=4932;			
FM	[1]			
FP	SEQUENCE FROM N.A.			
RX	MEDLINE=87286864; PubMed=3302672;			
RA	Trueheart J., Boeke J.D., Fink G.R.;			
RT	"Two genes required for cell fusion during			
RT	yeast cell fusion: evidence			
RT	for a pheromone-induced surface protein."			
RL	Mol. Cell. Biol. 7:2316-2328(1987).			
DR	EMBL: M16717; AAA34615.1; ..			
DR	SGD; S0000533; RNO1.			
40	SEQUENCE 407 AA: 43262 MW; 682694.0E74A804C3 CRC64;			

Query Match	22.8%	Score 160	DB 3	Length 407
Best Local Similarity	38.1%	Pred. No.	3.7e-08	
Matches	51	Conservative	5	Mismatches 34
				Indels 44
				Gaps 8

[illegible]

Search completed: February 1, 2002, 15:07:10
Job time: 315 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 1, 2002, 15:02:20 ; Search time 15.38 seconds
(without alignments)
293.223 Million cell updates/sec

Title: US-09-591-632-2_COPY_1_123

Perfect score: 701
Sequence: 1 MSDSNOGNNQNYQYQSQNG.....FNYNNNLQYQAGFQPSQSG 123

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	701	100.0	685	1	ERF2_YEAST
2	228	32.5	715	1	ERF2_CANAL
3	167	23.8	592	1	R856_HUMAN
4	166	23.7	405	1	YCC8_YEAST
5	166	23.7	741	1	ERF2_PICPT
6	157	22.4	512	1	FUS_BOVIN
7	147	21.0	462	1	ANK7_DICDI
8	143	20.4	372	1	P65H_MYCGE
9	143	20.4	518	1	FUS_MOUSE
10	141	20.1	387	1	SSXT_MOUSE
11	139	19.8	418	1	SSXT_MOUSE
12	136	19.4	526	1	FUS_HUMAN
13	135	19.3	128	1	YBM6_YEAST
14	132	18.8	882	1	IF2_BORBU
15	128	18.3	353	1	ROD_RAT
16	127	18.1	355	1	ROD_HUMAN
17	126.5	18.0	738	1	YKRF_YEAST
18	123	17.5	346	1	YNS3_CAEEL
19	122.5	17.5	750	1	CBMB_BACTV
20	121	17.3	954	1	XYNA_RUMFL
21	118	16.8	838	1	GLT4_WHEAT
22	116	16.5	298	1	3AKD_MYCPA
23	115.5	16.5	826	1	CRNA_BACUH
24	115	16.4	839	1	GLT5_WHEAT
25	115	16.4	991	1	DHPL_SCHPO
26	114.5	16.3	552	1	ERF1_CANAL
27	114	16.3	295	1	LEG3_CANFA
28	113.5	16.2	307	1	SYPH_BOVIN
29	113	16.1	753	1	CKAA_BACUP
30	112	16.0	460	1	YSAL_CAEEL
31	112	16.0	1028	1	OVO_DROME
32	111.5	15.9	313	1	SYPH_HUMAN
33	111	15.8	365	1	ROAL_DROME

34	111	15.8	870	1	YSO5_CAEEL	Q10130 caenorhabd
35	110.5	15.8	346	1	RO21_XENLA	P51989 xenopus lae
36	110.5	15.8	1902	1	SMF1_HUMAN	O14497 homo sapien
37	109.5	15.6	159	1	CUP9_DROME	P27781 drosophila
38	107.5	15.3	1048	1	ACOL_ARATH	O04379 arabidopsis
39	107	15.3	401	1	YK03_CAEEL	P34291 caenorhabd
40	107	15.3	1231	1	YK73_CAEEL	P34314 caenorhabd
41	106.5	15.2	660	1	GLT3_WHEAT	P08478 triticum ae
42	105.5	15.0	345	1	SOD_DROME	O08473 drosophila
43	105.5	15.0	648	1	GLT0_WHEAT	P10387 triticum ae
44	104.5	14.9	159	1	GRPA_MEDPA	O09134 medicago fa
45	104.5	14.9	655	1	EMS_MOUSE	O61545 mus musculu

ALIGNMENTS

RESULT ID	ERF2_YEAST	STANDARD	PRT	685 AA.
AC	P05453; P05420;			
DT	01-NOV-1988 (rel. 09, Created)			
DT	01-NOV-1988 (rel. 36, Last annotation update)			
DE	EUKARYOTIC PEPTIDE CHAIN RELEASE FACTOR GTP-BINDING SUBUNIT (ERF2)			
DE	(TRANSLATION RELEASE FACTOR 3) (ERF3) (ERF-3) (OMNIPOTENT SUPPRESSOR			
DE	PROTEIN 2) (G1 TO S PHASE TRANSITION PROTEIN 1)).			
GN	SUP35 OR SUP2 OR SUP12 OR GST1 OR SAL3 OR PNM2 OR YDR172W OR			
GN	YD9395.05.			
OS	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; Saccharomycetaceae; Saccharomyces.			
OX	NCBI_TaxID=4932;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=88329727; PubMed=3047009;			
RA	Kushnirov V.V., Ter-Avanesyan M.D., Telckov M.V., Surguchov A.P.,			
RA	Smirnov V.N., Inge-Vechtomov S.G.;			
RT	"Nucleotide sequence of the SUP2 (SUP35) gene of Saccharomyces			
RT	cerevisiae.";			
RL	Gene 66:45-54(1988).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=87219095; PubMed=3556215;			
RA	Kushnirov V.V., Ter-Avanesyan M.D., Surguchov A.P., Smirnov V.N.,			
RA	Inge-Vechtomov S.G.;			
RT	"Localization of possible functional domains in sup2 gene product of			
RT	the yeast Saccharomyces cerevisiae.";			
RL	FEBS Lett. 215:257-260(1987).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=88172503; PubMed=3280807;			
RA	Wilson P.G., Culbertson M.R.;			
RA	"SUP2 suppressor protein of yeast. A fusion protein related to the			
RT	EF-1 family of elongation factors.";			
RL	J. Mol. Biol. 199:559-573(1988).			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=88296422; PubMed=2841115;			
RA	Kukuchi Y., Shimatake H., Kikuchi A.;			
RA	"A yeast gene required for the G1-to-S transition encodes a protein			
RT	containing an A-kinase target site and GTPase domain.";			
RL	EMBO J. 7:1175-1182(1988).			
RN	[5]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=S288C / AB972;			
RA	Murphy L., Harris D.E., Barrell B.G., Rajandream M.A.;			
RA	Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.			
RN	[6]			
RP	FUNCTION.			
RX	MEDLINE=96016209; PubMed=7556078;			
RA	Stanfield I., Jones K.M., Kushnirov V.V., Dagkasmanakaya A.R.,			
RA	Poznyakovskii A.I., Paushtin S.V., Ntteras C.R., Cox B.S.,			

RL Genomics 38:51-57(1996).
RN [2]
RP SEQUENCE FROM N.A. (SHORT ISOFORM), AND SEQUENCE OF 282-297 & 307-320.
RA MEDLINE-97045110; PubMed-8890175;
RX Bertolotti A., Lutz Y., Heard D.J., Chambon P., Tora L.;
RT "hTF(II)88, a novel RNA/ssDNA-binding protein with homology to the
RT pro-oncoproteins TLS/FUS and EMS is associated with both TFIID and RNA
RT polymerase II.";
RL Embo J. 15:5022-5031(1996).
RN [3]
RP SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS).
RX MEDLINE-99013873; PubMed-9795213;
RA Morohoshi F., Ootsuka Y., Arai K., Ichikawa H., Mitani S.,
RA Munakata M., Ohki M.;
RT "Genomic structure of the human RBP56/hTFII68 and FUS/TLS genes.";
RL Gene 221:191-198(1998).
CC -1- FUNCTION: RNA AND SSDNA-BINDING PROTEIN THAT MAY PLAY SPECIFIC
CC ROLES DURING TRANSCRIPTION INITIATION AT DISTINCT PROMOTERS. CAN
CC ENTER THE PREINITIATION COMPLEX TOGETHER WITH THE RNA POLYMERASE
CC II (POL II).
CC -1- SUBUNIT: BELONGS TO THE RNA POLYMERASE II (POL II) TRANSCRIPTIONAL
CC MULTIPROTEIN COMPLEX, TOGETHER WITH THE TATA-BINDING PROTEIN (TBP)
CC AND OTHER TBP-ASSOCIATED FACTORS (TAF(II)S).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG ISOFORM (SHOWN HERE)
CC AND A SHORT ISOFORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: UBIQUITOUS. OBSERVED IN ALL FETAL AND ADULT
CC TISSUES.
CC -1- DISEASE: A FORM OF EXTRASKELETAL MYXOID CHONDROSARCOMAS (EMC) IS
CC CHARACTERIZED BY A CHROMOSOMAL TRANSLOCATION T(9;17)(Q22,Q11) THAT
CC INVOLVES TAF2N AND NR4A3.
CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
CC -1- SIMILARITY: CONTAINS 1 RANBP-TYPE ZINC FINGER.
CC -1- SIMILARITY: BELONGS TO THE TET FAMILY OF RNP PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U51334; AAC50932.1; -
DR EMBL: X98893; CA67398.1; -
DR EMBL: AB010067; BAA33811.1; -
DR EMBL: AB010057; BAA33811.1; JOINED.
DR EMBL: AB010058; BAA33811.1; JOINED.
DR EMBL: AB010059; BAA33811.1; JOINED.
DR EMBL: AB010060; BAA33811.1; JOINED.
DR EMBL: AB010061; BAA33811.1; JOINED.
DR EMBL: AB010062; BAA33811.1; JOINED.
DR EMBL: AB010063; BAA33811.1; JOINED.
DR EMBL: AB010064; BAA33811.1; JOINED.
DR EMBL: AB010065; BAA33811.1; JOINED.
DR EMBL: AB010066; BAA33811.1; JOINED.
DR EMBL: AB010067; BAA33812.1; -
DR EMBL: AB010067; BAA33812.1; -
DR EMBL: AB010057; BAA33812.1; JOINED.
DR EMBL: AB010058; BAA33812.1; JOINED.
DR EMBL: AB010059; BAA33812.1; JOINED.
DR EMBL: AB010060; BAA33812.1; JOINED.
DR EMBL: AB010061; BAA33812.1; JOINED.
DR EMBL: AB010062; BAA33812.1; JOINED.
DR EMBL: AB010063; BAA33812.1; JOINED.
DR EMBL: AB010064; BAA33812.1; JOINED.
DR EMBL: AB010065; BAA33812.1; JOINED.
DR EMBL: AB010066; BAA33812.1; JOINED.
DR MIM: 601574; -
DR InterPro: IPR000504; RRM.
DR InterPro: IPR001876; Znf-RanBP.
DR Pfam: PF00076; rrm; 1.
DR Pfam: PF00641; zf-RanBP; 1.
DR SMART: SM00360; RRM; 1.

DR SMART: SM00547; Znf_RB2; 1.
DR PROSITE: PS50102; RRM; 1.
DR PROSITE: PS00030; RRM_RNP_1; FALSE_NEG.
KW Proto-oncogene; RNA-binding; DNA-binding; Nuclear protein; Repeat;
KW Alternative splicing; Chromosomal translocation; Zinc-finger; Zinc;
KW Metal-binding.
FT DOMAIN 1 208 GIN/GLY/SER/TYR-RICH.
FT DOMAIN 234 320 RNA-BINDING (RRM).
FT DOMAIN 320 350 ARG/GLY-RICH.
FT ZN_FING 360 379 C4-TYPE (POTENTIAL).
FT DOMAIN 407 575 21 X APPROXIMATE/TANDEM REPEATS OF D-R-
FT [S,G](0,3)-G-G-Y-G-G.
FT REPEAT 407 413 1.
FT REPEAT 414 420 2.
FT REPEAT 421 429 3.
FT REPEAT 430 439 4.
FT REPEAT 440 448 5.
FT REPEAT 449 457 6.
FT REPEAT 458 465 7.
FT REPEAT 466 473 8.
FT REPEAT 474 481 9.
FT REPEAT 482 488 10.
FT REPEAT 489 496 11.
FT REPEAT 497 503 12.
FT REPEAT 504 510 13.
FT REPEAT 511 517 14.
FT REPEAT 518 524 15.
FT REPEAT 525 533 16.
FT REPEAT 534 543 17.
FT REPEAT 544 551 18.
FT REPEAT 552 560 19.
FT REPEAT 561 568 20.
FT REPEAT 569 575 21.
FT VARSPLIC 60 62 MISSING (IN SHORT ISOFORM).
SQ SEQUENCE 592 AA; 61830 MW; 73D37C171E1E2BCA CRC64;

Query Match 23.8%; Score 167; DB 1; Length 592;
Best Local Similarity 29.1%; Pred. No. 8; 8e-07;
Matches 46; Conservative 17; Mismatches 45; Indels 50; Gaps 7;

QY 1 MSDS-----NCGNNQONQOYSONGNOQGG-----NNRY--QGYOAYNAQAP 42
DB 1 MSDSGSGYGSGGEGQDQSTYTGKPNQSGYGOASQSYSGYGTOTDSSYQONYSGYSYQSQ 60
QY 43 GGYOAYNQGYSGYOQGGYQO--YN-----PDAGYQOQYNPQG 77
DB 61 SGYSQSGYSGYENQKQSSYSGQPYNNQGOQONMESSGSGGRAPSDQPDYGGQDSDYQGS 120
QY 78 GYQOYNPQGYQOQFNPQGRGKYNKNFNNNNLOGYQA 115
DB 121 GYDQH--QGSYDQSN-----YDQHDYSQONQSYHS 151

RESULT 4
YC8_YEAST
ID YC8_YEAST STANDARD; PRT; 405 AA.
AC P25367;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE HYPOTHETICAL 42.6 KDA PROTEIN IN BIK1-FUS1 INTERGENIC REGION.
GN YCL028W OR YCL28M OR YCL181.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-91377317; PubMed-1897318;
RA Rad M.R., Luetzenkirchen K., Xu G., Kleinhaus U., Hollenberg C.P.;
RT "The complete sequence of a 11,953 bp fragment from ClG on chromosome
III encompasses four new open reading frames.";


```

CC -1- FUNCTION: BINDS BOTH SINGLE-STRANDED AND DOUBLE-STRANDED DNA AND
CC PROMOTES ATP-INDEPENDENT ANNEALING OF COMPLEMENTARY SINGLE-
CC STRANDED DNAS AND D-LOOP FORMATION IN SUPERHELICAL DOUBLE-STRANDED
CC DNA. MAY PLAY A ROLE IN MAINTENANCE OF GENOMIC INTEGRITY (BY
CC SIMILARITY).
CC -1- SUBUNIT: COMPONENT OF NUCLEAR RIBOPROTEIN COMPLEXES (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR, EXHIBITS DIFFUSE STAINING
CC THROUGHOUT (EXCLUDING NUCLEOLI), TOGETHER WITH A SMALL NUMBER OF
CC INTENSELY STAINED FOCAL POINTS, OR GRANULES, AND PUNCTATE STAINING
CC ALONG THE NUCLEAR ENVELOPE.
CC -1- DOMAIN: THE C-TERMINAL DOMAIN BINDS CARBOHYDRATES.
CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
CC -1- SIMILARITY: BELONGS TO THE TET FAMILY OF RNP PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL: U26024; AAC13543.1; -
CC HSP: P09651; 1HA1.
CC InterPro: IPR000504; RRM.
CC InterPro: IPR01876; Znf-RanBP.
CC Pfam: PF00076; rrm. 1.
CC DR Pfam: PF00641; zf-RanBP. 1.
CC DR SMART: SM00360; RRM; 1.
CC DR SMART: SM00547; Znf_RBZ; 1.
CC DR PROSITE: PS50102; RRM; 1.
CC DR RNA-binding: P500030; RRM_RNP_1; FALSE_NEG.
CC DR RNA-binding: DNA-binding: Nuclear protein; Repeat; zinc-finger; zinc:
CC Metal-binding.
CC KM DOMAIN 1 164 GLN/GLY/SER/TYR-RICH.
CC FT DOMAIN 165 253 GLY-RICH.
CC FT DOMAIN 271 357 RNA-BINDING (RRM).
CC FT DOMAIN 357 512 ARG/GLY-RICH.
CC FT ZN_FING 414 433 C4-TYPE (POTENTIAL).
CC FT ZN_FING 414 433 C4-TYPE (POTENTIAL).
CC SO SEQUENCE 512 AA: 52240 MW: 3652329C044F1386 CRC64;

Query Match 22.4%; Score 157; DB 1; Length 512;
Best Local Similarity 28.9%; Pred. No. 4.9e-06;
Matches 44; Conservative 12; Mismatches 62; Indels 34; Gaps 5;

QY 2 SBSNOCN--NQONTQOY SQNGNQOQGNRRYQYQATNAAQAPAGGYQNTQYSGYQQ-- 57
DB 84 SOSQSQSSYQSSSPCYGQDPAPPSSTSGSSSSQSSGYGQPOGGGQGGGQCGQOQSY 143
QY 58 GGYYQYNPDAGY--QOQYNYNQ-----GGYYQYNPDAGYQOQ 91
DB 144 GQOQOQYNPDAGYQOQOQYNYSSGGGGGGGGGGGYQOQDPMSMSGGGGGGGGGGGG 203
QY 92 FNPQGRGKYNKFNKNNNNLQGYQAGFQPOQSG 123
DB 204 QQDRGGRGKGGGGYNNRS-----SGGYEPRGRG 231

RESULT 7
ANX7_DICDI STANDARD; PRT; 462 AA.
AC P24639;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ANNEXIN A7 (ANNEXIN VII) (SYNEXIN).
GN NXNA OR ANN7.
OS Dictyostelium discoideum (slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN-AX2;
RX MEDLINE-91198150; PubMed-1826615;
RA Greenwood M., Tsang A.;
RT "Sequence and expression of annexin VII of Dictyostelium discoideum.";
RL Biochim. Biophys. Acta 1088:429-432(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-91373375; PubMed-1832674;
RA Döring V., Schleicher M., Noegel A.A.;
RT "Dictyostelium annexin VII (synexin). cDNA sequence and isolation of
RT a gene disruption mutant.";
RL J. Biol. Chem. 266:17509-17515(1991).
RN [3]
RP PARTIAL SEQUENCE OF 143-437.
RX MEDLINE-91107669; PubMed-1824843;
RA Gerke V.;
RT "Identification of a homologue for annexin VII (synexin) in
RT Dictyostelium discoideum.";
RL J. Biol. Chem. 266:1697-1700(1991).
RN [4]
RX X-RAY CRYSTALLOGRAPHY (2.45 ANGSTROMS) OF 108-462.
RA Liemann S., Bringewiler I., Benz J., Goettig P., Hofmann A., Huber R.,
RA Noegel A.A., Jacob U.;
RT "Crystal structure of the C-terminal tetrad repeat from synexin
RT (annexin VII) of Dictyostelium discoideum.";
RL J. Mol. Biol. 270:79-88(1997).
CC -1- FUNCTION: CALCIUM/PHOSPHOLIPID-BINDING PROTEIN WHICH PROMOTES
CC MEMBRANE FUSION AND IS INVOLVED IN EXOCYTOSIS.
CC -1- SIMILARITY: BELONGS TO THE ANNEXIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL: X60269; CAA42815.1; -
CC DR EMBL: X60270; CAA42815.1; JOINED.
CC DR EMBL: X60270; CAA42815.1; -
CC DR EMBL: M69022; AAA33166.1; -
CC DR PIR: S14723; LUD07.
CC DR HSP: P26256; 1AE1.
CC DR DictyDb: DD01033; nxna.
CC DR InterPro: IPR001464; Annexin.
CC DR Pfam: PF00191; annexin. 4.
CC DR SMART: SM00335; ANX; 4.
CC DR PROSITE: PS00223; ANNEXIN; 2.
CC KW Annexin; Calcium/phospholipid-binding; Repeat; Alternative splicing.
CC FT DOMAIN 1 131 ANNEXIN 1.
CC FT REPEAT 170 230 ANNEXIN 1.
CC FT REPEAT 242 302 ANNEXIN 2.
CC FT REPEAT 324 384 ANNEXIN 3.
CC FT REPEAT 401 461 ANNEXIN 4.
CC FT VARSPPLIC 13 55 MISSING (IN SHORT ISOFORM).
CC FT CONFLICT 245 245 A -> R (IN REF. 2).
CC SO SEQUENCE 462 AA: 51172 MW: 6C846E480A69AF9 CRC64;

Query Match 21.0%; Score 147; DB 1; Length 462;
Best Local Similarity 40.7%; Pred. No. 2.8e-05;
Matches 55; Conservative 2; Mismatches 40; Indels 38; Gaps 11;

QY 5 NQG--NQONTQOY SQNGNQOQGNRRYQY---QAYNAQAPAGGY--YQNYGYSGY-QQ 57
DB 6 NQGYPPQSNSPQPGQYCAPQGGYPPQGGYPPQGGY-----PPQGGYPPQGGYPPQGGYPPQ 61
QY 58 GGY---QOYNPADAGY--QOQYNYNQGGY---QOYNPQGGY--QOQFNPQGRGKYNKFN 107
DB 62 QGYPPQGGYPPQGGYPPQGGYPPQGGYPPQGGYPPQGGYPPQGGYPPQGGYPPQGGYPPQ 109

```

OY 108 NNLOQY--QAGFOPQ 120
 DB 110 ---QGYPPQOGGYPO 121

RESULT 8

658L_MYCGE STANDARD; PRT; 372 AA.

DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PROLINE-RICH P65 PROTEIN HOMOLOG.

OS Mycoplasma genitalium.
 OC Bacteria; Firmicutes; Bacilllus/Clostridium group; Mollicutes;
 OC Mycoplasmataceae; Mycoplasma.
 OX NCBI_TaxID=2097;

SEQUENCE FROM N.A.

RP STRAIN-ATCC 33530 / G-37;
 RX MEDLINE=96026346; PubMed=7569993;

RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
 RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
 RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhmann J.L.,
 RA Nguyen D.T., Uterback T.R., Sauder D.M., Phillips C.A., Merrick J.M.,
 RA Tomb J.F., Dougherty B.A., Bort K.F., Hu P.-C., Lucier T.S.,
 RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.,
 RT "The minimal gene complement of Mycoplasma genitalium";
 RL Science 270:397-403(1995).

-1- SIMILARITY: TO M.PNEUMONIAE PROTEIN P65.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: U39701; AAC71436.1; -

DR TIGR: MG217; -

KW Repeat; Complete proteome.

FT DOMAIN 29 128 8 x DPNOQ(O)FNQ REPEATS.

FT REPEAT 29 37 1.

FT REPEAT 41 49 2.

FT REPEAT 61 69 3.

FT REPEAT 81 89 4.

FT REPEAT 101 109 5.

FT REPEAT 120 128 6.

FT REPEAT 139 147 7.

FT REPEAT 120 128 8.

SEQUENCE 372 AA; 44664 MW; 4C29701D213CE19E CRC64;

Query Match 20.4%; Score 143; DB 1; Length 372;
 Best Local Similarity 32.7%; Pred. No. 4.8e-05;
 Matches 51; Conservative 12; Mismatches 55; Indels 38; Gaps 8;

OY 3 DSNOGNNOQYQYSONG---NOQGN-----RYQGYQAYNMAQAPAGGYQNYQYSG 54
 DB 22 DNNQYQYQDPNQOQFNQSGFDPNQOQFNQSGFDPNQOQFNQAGFDPNQOQYQDP 81
 OY 55 ---YQGGYQYQYVNPAGY-QQOYVNPAGYQYQYVNPAGYQYQYQYQYQYQYQYQY 94
 DB 82 PNOQOQFNQGFDPNQOQYQYQDPNQOQFN-QAGFDPNQOQYQYQDPNQOQFNQSGFDPNQOQYQDP 140
 OY 95 QGGRGNKYNFNNNNNLOQYQAGFQ-----PQSQ 122
 DB 141 NQOQFNQPSFDLNNQ-QFNQGFNQSAPFRTQEQ 175

RESULT 9
 FUS_MOUSE STANDARD; PRT; 518 AA.

AC P56959;

DT 20-AUG-2001 (Rel. 40, Created)

DT 20-AUG-2001 (Rel. 40, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE RNA-BINDING PROTEIN FUS (PIGEPEN PROTEIN).

FUS.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

SEQUENCE FROM N.A.

RA Alapet S.R., Zhang M., Zhao X., Allegro M.A., Allegro M.C.,

RA Burdall C.A.;

RT "Regulation of p19pen expression in mouse embryos";

RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: BINDS BOTH SINGLE-STRANDED AND DOUBLE-STRANDED DNA AND

PROMOTES ATP-INDEPENDENT ANNEALING OF COMPLEMENTARY SINGLE-

STRANDED DNAs AND D-LOOP FORMATION IN SUPRHELICAL DOUBLE-STRANDED

DNA. MAY PLAY A ROLE IN MAINTENANCE OF GENOMIC INTEGRITY (BY

SIMILARITY).

CC -1- SUBUNIT: COMPONENT OF NUCLEAR RIBOPROTEIN COMPLEXES (BY

SIMILARITY).

CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).

CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).

CC -1- SIMILARITY: BELONGS TO THE TET FAMILY OF RNP PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: AF224264; AAF70602.1; -

DR MGD: MGI:1353633; FUS.

DR InterPro: IPR002952; Eggshell.

DR InterPro: IPR000504; RRM.

DR InterPro: IPR001876; Znf-RanBP.

DR Pfam: PF00076; rrm; 1.

DR Pfam: PF00641; zf-RanBP; 1.

DR SMART: SM00360; RRM; 1.

DR SMART: SM00547; Znf-RB2; 1.

DR PROSITE: PS0102; RRM; 1.

DR PROSITE: PS00030; RRM_RNP_1; FALSE_NEG.

DR RNA-binding; DNA-binding; Nuclear protein; Repeat; Zinc-finger; Zinc;

KW Metal-binding.

FT DOMAIN 1 167 GLN/GLY/SER/TYR-RICH.

FT DOMAIN 168 265 GLY-RICH.

FT DOMAIN 278 364 RNA-BINDING (RRM).

FT DOMAIN 364 518 ARG/GLY-RICH.

FT ZN_FING 421 440 C4-TYPE (POTENTIAL).

SEQUENCE 518 AA; 52673 MW; E06F231BFED78D6 CRC64;

Query Match 20.4%; Score 143; DB 1; Length 518;
 Best Local Similarity 29.1%; Pred. No. 6.5e-05;
 Matches 46; Conservative 12; Mismatches 60; Indels 40; Gaps 7;

OY 2 SDSNOGN--NOQNYQYSONGNOQGNRRQGYDAYNAQAPAGGYQNYQYSGYSGY 58
 DB 85 SOSOSSSTYGQSSSTPGGQAPSPSTSGSYGSSOSSSTYGQPOSGYGGQSSGTYGQSSY 144
 OY 59 GYQYQ--YNPAGY--QQOYVNPQ-----GQY-----QOYVNPQ 85
 DB 145 GQOQSSSYNPNPGYQY 204
 OY 86 GGYQOQFNQPGGGRGNKYNFNNNNNLOQYQAGFQPOSQ 123

Matches 52; Conservative 12; Mismatches 47; Indels 30; Gaps 12;

OY 6 OGNNOONV--QOXSONG-----NOO--QGNRRYGOAYNAQAPAGGYONOGYS 53
 DB 263 QYSGQEDYGDQYSHGQGPPEGMNOQYYPDGHNDY-GYV---QPSYPEQGYDPEYDSS 318
 OY 54 -GYOQGGYOOYNDPAGYOOQYNDPQGY---QOYNDPQGY---QOQFNP-QGGRG-NYKN 103
 DB 319 OHYEGGNSGNGYGGQODAYGQPPQGGYPPQGGYPPQGGYPPQGGYPPQGGYPPQGGYPPN 378
 OY 104 FNNNNLQGYQ-AGFQPGQSG 123
 DB 379 YPG---QGOYGGYRPTPG 396

RESULT 12
 FUS_HUMAN
 ID FUS_HUMAN STANDARD: PRT; 526 AA.
 AC P35637;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE RNA-BINDING PROTEIN FUS (ONCOGENE FUS) (ONCOGENE TLS) (TRANSLATED IN LIPOSARCOMA PROTEIN) (POM75) (75 KDA DNA-PAIRING PROTEIN).
 DE LIPOSARCOMA PROTEIN (POM75) (75 KDA DNA-PAIRING PROTEIN).
 GN FUS OR TLS.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (LONG ISOFORM).
 RX MEDLINE=93288139; PubMed=8510758;
 RA Crozat A., Aman P., Mandahl N., Ron D.;
 RT "Fusion of CHOP to a novel RNA-binding protein in human myeloid liposarcoma".
 RL Nature 363:640-644(1993).
 RN [2]
 RP SEQUENCE FROM N.A. (SHORT ISOFORM).
 RX MEDLINE=93350637; PubMed=7503811;
 RA Rablitt T.H., Forster A., Larson R., Nathan P.;
 RT "Fusion of the dominant negative transcription regulator CHOP with a novel gene FUS by translocation t(12;16) in malignant liposarcoma".
 RL Nat. Genet. 4:175-180(1993).
 RN [3]
 RP SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS).
 RX MEDLINE=99013873; PubMed=9795213;
 RA Morohoshi F., Ootsuka Y., Arai K., Ichikawa H., Micani S.,
 RT "Genomic structure of the human RBP56/htraf168 and FUS/TLS genes".
 RL Gene 221:191-198(1998).
 RN [4]
 RP SEQUENCE OF 235-244; 307-312; 335-345 AND 349-357, 6 CHARACTERIZATION.
 RX MEDLINE=20036580; PubMed=10567410;
 RA Baechtold H., Kuroda M., Sok J., Ron D., Lopez B.S., Akhmedov A.T.;
 RT "Human 75-kDa DNA-pairing protein is identical to the pro-oncoprotein TLS/FUS and is able to promote D-loop formation".
 RL J. Biol. Chem. 274:34337-34342(1999).
 RN [5]
 RP SEQUENCE OF 265-276; 317-331 AND 335-357, AND IDENTIFICATION.
 RX MEDLINE=99369251; PubMed=10442642;
 RA Bertrand P., Akhmedov A.T., Delacote F., Durrbach A., Lopez B.S.;
 RT "Human POM75 is identified as the pro-oncogene TLS/FUS: both POM75 and POM100 DNA homologous pairing activities are associated to cell proliferation".
 RL Oncogene 18:4515-4521(1999).
 RN [6]
 RP CHROMOSOMAL TRANSLOCATION.
 RX MEDLINE=94243799; PubMed=8187069;
 RA Ichikawa H., Shimizu K., Hayashi Y., Ohki M.;
 RT "An RNA-binding protein gene, TLS/FUS, is fused to ERG in human myeloid leukemia with t(16;21) chromosomal translocation".
 RL Cancer Res. 54:2865-2868(1994).
 CC -1- FUNCTION: BINDS BOTH SINGLE-STRANDED AND DOUBLE-STRANDED DNA AND

PROMOTES ATP-INDEPENDENT ANNEALING OF COMPLEMENTARY SINGLE-STRANDED DNAS AND D-LOOP FORMATION IN SUPERHELICAL DOUBLE-STRANDED DNA. MAY PLAY A ROLE IN MAINTENANCE OF GENOMIC INTEGRITY.
 CC -1- SUBUNIT: COMPONENT OF NUCLEAR RIBOPROTEIN COMPLEXES.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG ISOFORM (SHOWN HERE) AND A SHORT ISOFORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: UBIQUITOUS.
 CC -1- DISEASE: A FORM OF MALIGNANT MYELOID LIPOSARCOMA IS CHARACTERIZED BY A CHROMOSOMAL TRANSLOCATION T(12;16)(Q13;P11) THAT INVOLVES FUS AND CHOP.
 CC -1- DISEASE: A FORM OF ACUTE MYELOID LEUKEMIA (AML) IS CHARACTERIZED BY A CHROMOSOMAL TRANSLOCATION T(16;21)(P11;Q22) THAT INVOLVES FUS AND ERG.
 CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
 CC -1- SIMILARITY: BELONGS TO THE TET FAMILY OF RNP PROTEINS.
 CC -1- SIMILARITY: CONTAINS 1 RANBP-TYPE ZINC FINGER.
 CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: S62140; AAB27102.1; -;
 DR EMBL: S62138; AAB27103.1; ALT_SEQ.
 DR EMBL: X71427; CAA50538.1; ALT_SEQ.
 DR EMBL: X71428; CAA50559.1; ALT_SEQ.
 DR EMBL: AF071213; AAC35285.1; -;
 DR EMBL: AF071213; AAC35284.1; -;
 DR PIR: S33799; S33799.
 DR HSSP: P09651; LHA1.
 DR MIM: 137070; -;
 DR MIM: 131900; -;
 DR InterPro: IPR000504; RRM.
 DR InterPro: IPR001876; Znf-RanBP.
 DR Pfam: PF00076; rrm.1.
 DR Pfam: PF00641; zf-RanBP.1.
 DR SMART: SM00360; RRM.1.
 DR SMART: SM00547; Znf-RBZ.1.
 DR PROSITE: PS0102; RRM.1.
 DR PROSITE: PS00030; RRM_RNP.1; FALSE_NEG.
 KW Proto-oncogene; RNA-binding; DNA-binding; Nuclear protein; Repeat; Alternative splicing; Chromosomal translocation; Zinc-finger; Zinc; Metal-binding.
 KW Metal-binding.
 FT DOMAIN 1 165 GLN/GLY/SER/TYR-RICH.
 FT DOMAIN 2 166 267 GLY-RICH.
 FT DOMAIN 3 285 371 RNA-BINDING (RRM).
 FT DOMAIN 4 371 526 ARG/GLY-RICH.
 FT ZN_FING 422 453 RANBP-TYPE.
 FT SITE 266 267 BREAKPOINT FOR TRANSLATION TO FORM FUS/TLS-CHOP ONCOGENE.
 FT VARSPIC 64 65 TG -> S (IN SHORT ISOFORM).
 FT CONFLICT 338 338 T -> N (IN REF. 4).
 SQ SEQUENCE 526 AA; 53426 MW; 88CB8E263B7905549 CRC64;

Query Match 19.4%; Score 136; DB 1; Length 526;
 Best Local Similarity 27.0%; Pred. No. 0.00024;
 Matches 51; Conservative 9; Mismatches 57; Indels 72; Gaps 10;

OY 4 SNGNNOONQOXSQNGNQ-QGNRRYQY---QAYNAQAP-----AGY----- 45
 DB 29 ISQPYGOQSYSGYSTDTSYGYSQSYSGYSGQNTGCTGQSTPGQYSGTGSGSSSQ 88
 OY 46 --YONVGYSGY-----QGGYOOYNDPAGYOOQYNDPQGG 78
 DB 89 SSTGGQSSYTGQQAAPSTSSGSSSSSSSGYSGQPPSYGGQDQSSYGGQ-- 146
 OY 79 YOOYNDPQGY--QOQFNPQ-----GGRGNY-----KFNYYNNLQ--GYO 114


```

OX NCB1_TaxID-10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Kidney;
RA Arao Y., Kikuchi A.;
RT "Differential expression of AUP1 isoforms in rat tissues.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DDbJ databases.
CC -1- FUNCTION: BINDS WITH HIGH AFFINITY TO RNA MOLECULES THAT CONTAIN
CC AU-RICH ELEMENTS (ARES) FOUND WITHIN THE 3'-PRIME UNTRANSLATED
CC REGIONS OF MANY PROTOGENES AND CYTOKINE MRNAs (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR; COMPONENT OF RIBONUCLEOSOMES (BY
CC SIMILARITY).
CC -1- ALTERNATIVE PRODUCTS: 4 ISOFORMS: 1/P45 (SHOWN HERE), 2/P42, 3/P40
CC AND 4/P37; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: CONTAINS 2 RNA RECOGNITION MOTIFS (RRM).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AB046615; BAB03465.1; -
DR EMBL: AB046616; BAB03466.1; -
DR EMBL: AB046617; BAB03467.1; -
DR EMBL: AB046618; BAB03468.1; -
DR InterPro: IPR000504; RRM.
DR Pfam: PF00076; Rrm; 2.
DR SMART: SM00360; RRM; 2.
DR PROSITE: PS00102; RRM; 2.
DR PROSITE: PS00030; RRM_RNP_1; 2.
KW Nuclear protein; RNA-binding; Ribonucleoprotein; Repeat;
KW Alternative splicing.
FT DOMAIN 11 45 ALA-RICH.
FT DOMAIN 95 177 RNA-BINDING (RRM) 1.
FT DOMAIN 180 259 RNA-BINDING (RRM) 2.
FT DOMAIN 268 345 GLY-RICH.
FT DOMAIN 292 330 TYR-RICH.
FT VARSPLIC 77 95 MISSING (IN ISOFORM 2 AND ISOFORM 4).
FT VARSPLIC 283 332 GPSQNMNGYSNYMNGYSGYNSQYGGYGGYDYGYS
FT YGCGDYSN -> D (IN ISOFORM 3 AND ISOFORM
FT 4).
SO SEQUENCE 353 AA; 38192 MW; 6190EEL006F07DCB CRC64;

```

Query Match 18.3%; Score 128; DB 1; Length 353;
 Best Local Similarity 38.5%; Pred. No. 0.00072;
 Matches 30; Conservative 5; Mismatches 33; Indels 10; Gaps 1;

```

OY 17 SONGNQOQGNRRYQGYQATYAAQAPAGGYQNTQGYSGYQGGYQOQYNPDAGYQOQYNPQ 76
DB 285 SQMWNQGYSNYMWNGY-----GSYGYNSQYGGYGGYDYGYSNYSGYGGYDYSNQ 334
OY 77 GGYQATNPQGYQOQFNP 94
DB 335 SGYGYVSRRGHQNSTYKP 352

```

Search completed: February 1, 2002, 15:07:32
 Job time: 312 sec